X-RAY CRYSTALLOGRAPHIC STUDIES OF ENZYMES FOUND IN THE BIOSYNTHETIC PATHWAYS OF THIAMIN AND CYSTEINE, AND THE PURINE NUCLEOSIDE PHOSPHORYLASE SALVAGE PATHWAY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Christopher Thomas Jurgenson

August 2008

X-RAY CRYSTALLOGRAPHIC STUDIES OF ENZYMES FOUND IN THE BIOSYNTHETIC PATHWAYS OF THIAMIN AND CYSTEINE, AND THE PURINE NUCLEOSIDE PHOSPHORYLASE SALVAGE PATHWAY

Christopher Thomas Jurgenson, Ph. D.

Cornell University 2008

This work encompasses the structure of four different proteins, all of which were solved by X-ray crystallography. The structure of thiazole synthase (Thi4) in *Saccharomyces cerevisiae* involves the catabolism of another cofactor, nicotinamide adenine dinucleotide (NAD), through a series of catalytic steps that are now understood in light of several structural and biochemical studies. The product of the Thi4 reaction is an adenylated thiazole carboxylate, which is an unprecedented molecule found in primary metabolism. Thi4 also has a flavin adenine dinucleotide (FAD) binding fold, yet does not bind FAD. This implies that an ancient predecessor of Thi4 was able to bind FAD and evolved to bind NAD as a substrate without a change in topology.

The biosynthesis of cysteine in *Mycobacterium tuberculosis* involves a sulfur transfer step that uses a thiocarboxylated protein as the sulfur source. The protein complex structure of the cysteine synthase CysM and sulfur carrier protein CysO gives insight into how this reaction occurs. CysO has a β-grasp fold and is structurally similar to ubiquitin. The enzyme CysM binds to CysO in a pyridoxal 5´-phosphate dependent reaction with O-acetylserine (OAS) as the substrate to generate a cysteine residue on the C-terminus of CysO. CysM binds CysO through a loop that is ordered when CysO is bound and disordered when CysO is absent. Although there are other examples of proteins that bind to a sulfur carrier protein with a β-grasp fold, they all

differ in their composition of surface electrostatics responsible for protein-protein interactions.

The structure of the enzyme purine nucleoside phosphorylase (PNP) from the thermophile *Thermotoga maritima* is an example of a protein that has the same overall fold as its mesophilic counterpart bovine PNP, yet is stable at considerably higher temperatures. The differences in thermal stability are seen in the fewer number of glycine and greater number of proline residues in TmPNP over bPNP. Also, the protein-protein interactions between monomers in the biological unit of TmPNP contain a greater number of hydrophobic residues, thereby leading tighter packing and greater thermostability.

BIOGRAPHICAL SKETCH

Christopher Thomas Jurgenson was born on October 12, 1977 to Thomas and Barbara Jurgenson at Porter Memorial Hospital in Valparaiso Indiana. Christopher spent his childhood in Valparaiso, and graduated from Valparaiso High School in 1995. He then enrolled at Purdue University – Calumet where he studied microbiology for one year. In 1996, he moved to Indiana University – Bloomington where he majored in Biochemistry. Christopher graduated with honors with a B.S. degree in Biochemistry in 1998. After graduation, he went to work at Merck and Co. Inc., for five years in the Analytical Research department of Merck Research Laboratories. During that time, he attended graduate school part-time at Seton Hall University and obtained a M.S. degree in Chemistry in 2002. In May of 2002, Christopher married Genevieve Ann Sorrento, and since has had two children. Breanna Josephine was born on May 17, 2003, and Joshua Christopher was born on December 16, 2004. Christopher left Merck in 2003 to attend Cornell University.

This work is dedicated to my loving wife Genevieve who bore with me through all those harrowing days and sleepless nights.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Steven E. Ealick for giving me several interesting projects and the freedom to advance them in what way I saw fit. I thank Dr. Tadhg P. Begley and his students Abhishek Chatterjee and Kristin Burns for being outstanding collaborators. I thank Leslie Kinsland for her help in manuscript preparation as well as helping keep the lab in order and well supplied. I thank the National Institutes of Health for funding my graduate education through the Chemistry/Biology Interface Training Grant. I thank the American Crystallographic Association for financial support in the form of travel grants, and for inviting me to give talks at the annual ACA meeting on two occasions. I thank the North Eastern Collaborative Access Team for providing excellent support and facilities for synchrotron data collection. I thank the Cornell High Energy Synchrotron Source for providing a convenient and reliable facility on the Cornell campus for data collection. I thank Dr. Cynthia Kinsland for her support of my many projects, and providing an opportunity for me to learn techniques in molecular biology. I thank Dr. Ashley Deacon at the Joint Center for Structural Genomics at the Stanford Linear Accelerator Center for the dataset that led to the structure of *Thermotoga maritima* purine nucleoside phosphorylase. Finally, I would like to thank Shridhar, Katie, Erika, You-Na, Ying, Paola, Simon, Tim, Jarrod, Debamita, Mariya, Jolita and Yang for being great labmates and making this whole experience a lot more fun that it would have been otherwise.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

HET, hydroxyethylthiazole

HMP, hydroxymethylpyrimidine

HMP-P, hydroxymethylpyrimidine phosphate

TPP, thiamin pyrophosphate

ADT, adenosine diphospho-5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid

NAD, nicotinamide adenine dinucleotide

FAD, flavin adenine dinucleotide

PLP, pyridoxal 5′-phosphate

DTT, 1,4-dithiothreitol

r.m.s.d., root-mean-square difference

CS, Cysteine synthase

OAS, O-acetylserine

OASS, O-acetylserine sulfhydrylase

PNP, purine nucleoside phosphorylase

TmPNP, *Thermotoga maritima* purine nucleoside phosphorylase

bPNP, bovine purine nucleoside phosphorylase

Guanine, GUN

9-deazahypoxanthine, HPA

Ribose 1-phosphate, R1P

CHAPTER ONE

INTRODUCTION

Section 1.1 General Comments on Structural Biology

Modern methods in structural biology have changed the field that was once a science only practiced by a handful of creative computer programmers, crystallographers and chemists. Research in structural biology using X-ray crystallography has been streamlined through investments in better beamlines and tools for doing day to day crystallography experiments (*1*). Crystal screening is done using kits instead of making solutions from scratch, and sophisticated robots are capable of assembling screens using minute amounts of protein. Third generation synchrotrons have made improvements in overall brilliance, allowing for shorter exposure times and higher quality data. Advances in microdiffractometry allow for data collection on crystals that would have been too small to be useful a decade ago.

A testament to the impact of structural biology can be seen in the Nobel Prize lectures given by many laureates that include protein structures. The 2007 Nobel Prize in Chemistry given to Roger Kornberg for his work on "The molecular basis of eukaryotic transcription" was awarded in part due to X-ray structures of RNA polymerase II in complex with other transcription factors (*2*). The 2003 Nobel Prize in Chemistry was awarded to both Roderick MacKinnon and Peter Agre for their X-ray crystal structures of potassium and water channels, respectively (*3*). In recent years, nearly every Nobel Prize awarded for work done in biochemistry, cellular and molecular biology has been supported, at least in part, by either X-ray or, to a lesser extent NMR, protein structures.

These contributions to biology are due to the accessibility of X-ray crystallography as a research tool to anybody who chooses to use it. Beamtime at national synchrotron sources is plentiful, and the support staff at beamlines provide the

technical expertise required for everything from data collection to structural solution. There exists a vast array of computer programs that assist in data reduction (*4*), phasing of anomalous data (*5-10*), molecular replacement (*11, 12*), model building (*13, 14*) and refinement (*12, 15*) that are freely available for academic use. Phasing of anomalous data has become easier with the use of selenomethionine derivitized proteins synthesized by methionine auxotrophic overexpression strains of *Escherichia coli* instead of laborious heavy metal atom screening. In the event that selenomethionine derivitized protein crystals cannot be grown, phasing using the sulfur atoms of cysteine and methionine in native protein crystals has been achieved at a wavelength of 1.54 Å using synchrotron radiation (*16-18*) and the copper Kα emission band from a rotating anode generator (*19*). A newer technique involving the use of a chromium X-ray source with a Kα emission band at 2.29 Å has also been used to generate anomalous datasets to solve protein structures through sulfur atom phasing (*20-22*).

However, the limitations of X-ray crystallography are the same today as they have always been. The key to good X-ray protein crystal diffraction is a high quality crystal. Poorly ordered crystals will not diffract to atomic resolution, and small crystals suffer radiation decay faster than large ones which can affect the ability to collect a complete dataset. There are many was to address these problems such as extensive crystallization screening, mutating putative surface residues to decrease entropic effects (*23*), truncating multidomain enzymes and chemical cross-linking. One of the most effective methods is to screen several different orthologs of a particular enzyme. Since the surface residues are the ones involved in crystal packing, and are the least conserved residues among orthologous proteins, screening orthologs can increase the chance of obtaining a high quality crystal.

The future of structural biology lies in solving structures of large protein and protein-nucleic acid complexes, as well as solving structures of other proteins that are recalcitrant to crystallization. Recent work on the X-ray crystal structure of the actin initiator Arp2/3 complex consists of seven different proteins that are required for this biochemical process (*24-26*). Large protein complexes like the ribosome (*27-30*), non-ribosomal peptide synthases, fatty acid synthases (*31, 32*) and polyketide synthases are examples of large proteins that are the underpinning of some of the most important molecular biology of the cell. Without atomic resolution structures, currently only attainable through X-ray crystallography, we will never have more than a glimpse into the underlying chemistry behind these molecular machines.

Section 1.2 Thiazole Biosynthesis

Though much progress has been made in discovering the biosynthesis, uptake, salvage and regulation of thiamin in prokaryotes, comparatively little is still known about how eukaryotes make thiamin (Figure 1.1A). Like in bacteria, each moiety of thiamin pyrophosphate is generated through a separate pathway, coupled by a single enzyme and phosphorylated by a variety of kinases. The thiazole ring is made by the enzyme Thi4 from glycine, cysteine and a five carbon sugar (*33-35*). Thi5 generates the HMP moiety using pyridoxine and histidine (*36*), and is subsequently phosporylated and pyrophosphorylated by Thi20 (*37*). The two rings are joined by Thi6 (*38*) which is also a thiamin phosphate pyrophosphorylase and a hydroxyethylthiazole kinase (*39*).

The most recent discoveries in eukaryotic thiamin biosynthesis have been in the production of the thiazole ring. The enzyme Thi4 in yeast – Thi1 in *Arabidopsis thaliana* – is a homooctomer 35 kDa (30 kDa for Thi1) in size (Figure 1.1B). Orthologs of Thi4 in *Fusarium oxysporum*, *F. solani* (Sti35) and *Aspergillus oryzae* (ThiA) are transcriptionally regulated by TPP binding to a THI-box riboswitch (*40*). Thi4 is a dual function enzyme involved with mitochondrial DNA repair (*41*) as well

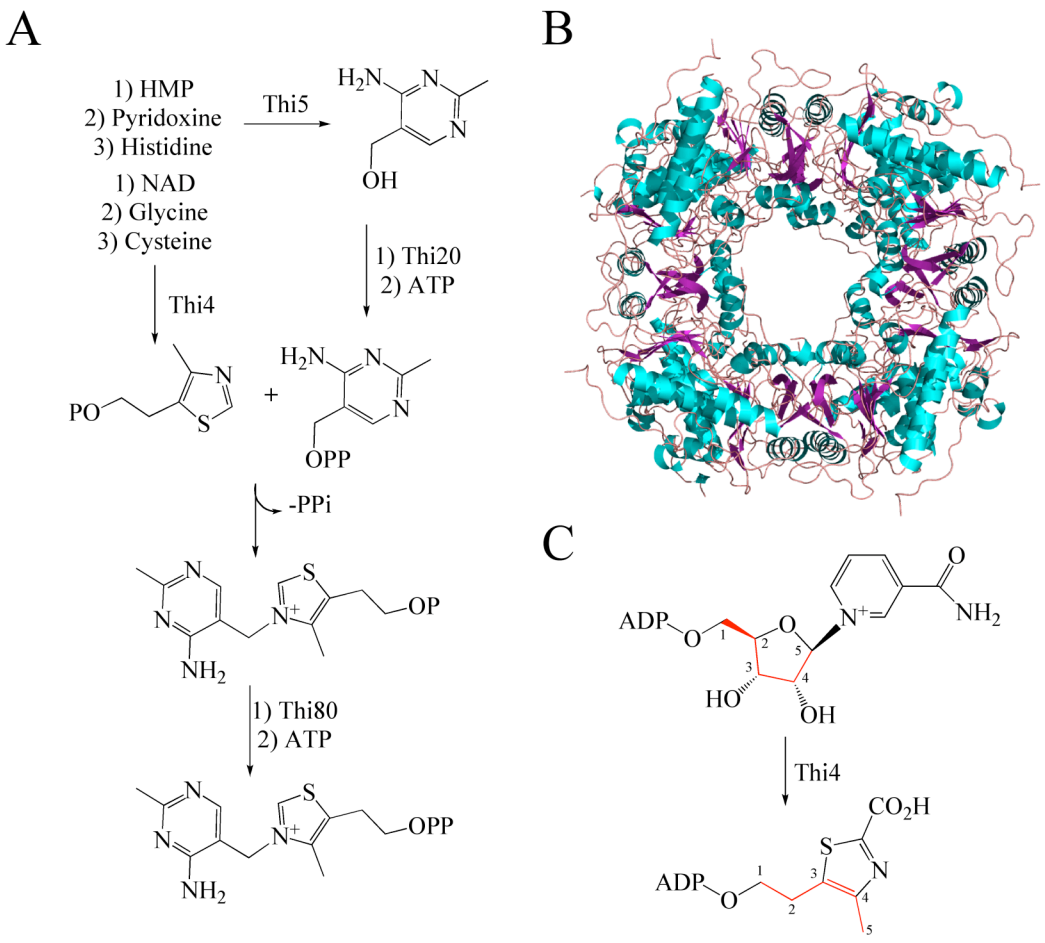


Figure 1.1: A) Overall thiamin biosynthetic pathway for *S. cerevisiae*. B) X-ray crystal structure of Thi4. C) Overall reaction catalyzed by Thi4. Atoms in the thiazole moiety that are similar to both molecules are labeled in red and numbered.

as thiazole biosynthesis. The discovery of a metabolite bound to Thi4 overexpressed in *E. coli* cells revealed an adenylated thiazole carboxylate (ADT) molecule through MS and NMR experiments (*42*). Crystal structures of Thi4 (*43*) and Thi1 (*44*) showed clear density for ADT, confirming the presence of this unprecedented metabolite. The adenylated ligand suggested that the substrate for Thi4 was a dinucleotide, but ADT was bound so tightly that it could not be removed from the native protein without denaturation. The structure of Thi4 was used to conduct structure activity relationship studies in order to generate a mutant that did not have endogenously bound ADT but still retained catalytic activity. C204S and H200N mutants produced an enzyme without a bound ligand yet were still biochemically active. Thi4 was incubated in the presence of NAD and FAD and was shown to produce four different compounds with NAD while no reactivity was observed with FAD. These compounds were identified as ADT, ADP-ribose, ADP-ribulose and an advanced intermediate which resulted from incubation of the C204A mutant of Thi4 with NAD and glycine (*45*). This represented the first example in a biosynthetic pathway where a co-factor is degraded to contribute to the molecular structure of another cofactor (*46*) (Figure 1.1C).

Section 1.3 Cysteine Biosynthesis

The biosynthesis of cysteine for most bacteria is done using O-acetylserine and sulfide as substrates for the pyridoxal 5′-phosphate (PLP)-dependent enzyme cysteine synthase (CS)/O-acetylserine sulfhydrylase (OASS) (*47*). The reaction is carried out through an α-aminoacrylate intermediate that serves as the electrophile for sulfide (*48*). Under normal growth conditions sulfide is relatively stable. However, the life cycle of *Mycobacterium tuberculosis* is more complex, as these bacteria possess several mechanisms allowing for their reproduction within the vacuole of macrophages. Some of these strategies include the ability to prevent the fusion of a phagosome containing the internalized bacterium with the host cell's lysosomal

system (*49*), the complex system of biologically active compounds which comprise the unique cell wall (*50-54*) and the expression of a catalase-peroxidase that protects the cell from oxidizing agents within the macrophage (*55*).

Another strategy is for the organism to eliminate the need for sulfate altogether in cysteine biosynthesis. In a highly oxidative environment, sulfide will oxidize to form a much less reactive sulfate ion, which is not a substrate for CS/OASS enzymes. A series of genes in the *M. tuberculosis* genome have been shown to be upregulated when bacilli are exposed to oxidative stress similar to the vacuole of a macrophage (*56*). Genes annotated as a cysteine synthase (CysM – Rv1336) a sulfur carry protein (CysO – Rv1335) and a zinc-hydrolase with a JAMM; Jab1/MPN domain metalloenzyme motif (Mec^+ – Rv1334) are responsible for this alternative pathway. CysO is posttranslationally modified to have a thiocarboxy C-terminus that serves as the sulfur source. This is done by another enzyme, MoeZ (Rv3206), which contains both adenyltransferase and rhodenase domains (*57*). CysO has a ubiquitin like fold, and is structurally similar to previously characterized sulfur carry proteins ThiS (*58*) and MoaD (*59*). CysM forms an α-aminoacrylate intermediate using O-acetylserine as the substrate like in other OASS enzymes (*57*). CysO binds to CysM with its thiocarboxy C-terminal tail pointing into the active site to deliver the sulfur atom. A CysO-cysteine adduct forms with a conventional peptide bond that is then cleaved through the hydrolytic activity of Mec^+ to release free cysteine.

Section 1.4 Purine Salvage

Purine nucleoside phosphorylase (PNP) enzymes are involved in the purine salvage pathway through reversible catalysis of purine nucleosides to generate a purine base and ribose 1-phosphate (*60*). This salvage pathway allows for the cell to metabolize preformed nucleotides into precursors for the synthesis of other nucleotides, thus bypassing the need for *de novo* nucleotide biosynthesis.

The fold and biological unit of bovine PNP (bPNP) (*61*) and *Thermotoga maritima* PNP (TmPNP) are the same, but each enzyme must function in a very different environment. The mesophilic enzyme bPNP is catalytically active and stable at a typical bovine calf internal temperature of 38.6°C, while TmPNP is active and stable at temperatures of up to 90°C (*62*). Some of the key structural differences between thermophiles and mesophiles are the number of proline residues, the presence of intramolecular disulfide bridges, number of salt bridges, hydrogen bonds, accessible surface area, increased polar surface area, hydrophibicity, fewer loop regions and greater compactness (*63, 64*). Many of these factors are interrelated, and it should be clear that strong intramolecular and intermolecular interactions, regardless of type, will ultimately lead to an increase in ΔG of protein unfolding. Comparison between TmPNP and bPNP show that these factors are present, and are largely the cause of their relative stabilities at high temperatures.

BIBLIOGRAPHY

(1) Dauter, Z. (2006) Current state and prospects of macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr 62*, 1-11.

(2) Kornberg, R. D. (2007) The molecular basis of eukaryotic transcription. *Prix Nobel*, 176-195.

(3) Agre, P., and MacKinnon, R. (2003) The Nobel Prize in Chemistry 2. *G.I.T. Laboratory Journal, Europe 7*, 294-295.

(4) Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307-326.

(5) Terwilliger, T. C. (2000) Maximum likelihood density modification. *Acta Crystallogr. D 56*, 965-972.

(6) Terwilliger, T. C., and Berendzen, J. (1999) Automated structure solution for MIR and MAD (http://www.solve.lanl.gov). *Acta Crystallogr. D 55*, 849-861.

(7) Hauptman, H. A. (1991) A minimal principle in the phase problem, in *Crystallographic computing 5: From Chemistry to Biology* (Moras, D., Podnarny, A. D., and Thierry, J. C., Eds.) pp 324-332, IUCr and Oxford University Press, Oxford, UK.

(8) Hauptman, H. (1990) A minimal principle in the direct methods of x-ray crystallography. *NATO ASI Series C315*, 3-15.

(9) Miller, R., and Weeks, C. M. (1998) Shake-and-Bake: applications and advances [for solving protein structures]. *NATO ASI Series C507*, 389-400.

(10) Weeks, C. M., and Miller, R. (1999) Optimizing Shake-and-Bake for proteins. *Acta Crystallogr. D55*, 492-500.

(11) Navaza, J., and Saludjian, P. (1997) AMoRe: an automated molecular replacement program package. *Methods Enzymol. 276*, 581-594.

(12) Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D 54*, 905-21.

(13) Terwilliger, T. C. (2002) Automated main-chain model-building by template-matching and iterative fragment extension. *Acta Crystallogr. D D59*, 34-44.

(14) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D 60*, 2126-32.

(15) Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography. *Acta Crystallogr. D 50*, 760-763.

(16) Dauter, Z., Dauter, M., De La Fortelle, E., Bricogne, G., and Sheldrick, G. M. (1999) Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J. Mol. Biol. 289*, 83-92.

(17) Agarwal, R., Bonanno, J. B., Burley, S. K., and Swaminathan, S. (2006) Structure determination of an FMN reductase from *Pseudomonas aeruginosa* PA01 using sulfur anomalous signal. *Acta Crystallogr. D D62*, 383-391.

(18) Uson, I., Schmidt, B., von Buelow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z., and Sheldrick, G. M. (2003) Locating the anomalous scatterer substructures in halide and sulfur phasing. *Acta Crystallogr. D D59*, 57-66.

(19) Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G., and Henriksen, A. (2004) Solving the structure of the bubble protein using the anomalous sulfur signal from single-crystal in-house Cu K–αλπηα diffraction data only. *Acta Crystallogr. D 60*, 250-5.

(20) Watanabe, N., Kitago, Y., Tanaka, I., Wang, J. W., Gu, Y. X., Zheng, C. D., and Fan, H. F. (2005) Comparison of phasing methods for sulfur-SAD using in-house chromium radiation: case studies for standard proteins and a 69 kDa protein. *Acta Crystallogr. D D61*, 1533-1540.

(21) Kitago, Y., Watanabe, N., and Tanaka, I. (2005) Structure determination of a novel protein by sulfur SAD using chromium radiation in combination with a new crystal-mounting method. *Acta Crystallogr. D D61*, 1013-1021.

(22) Rose, J. P., Liu, Z.-J., Temple, W., Chen, L., Lee, D., Newton, M. G., and Wang, B.-C. (2004) Practical aspects of SAS structure determination using chromium X-rays. *Rigaku Journal 21*, 3-11.

(23) Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr. D 63*, 636-45.

(24) Pollard, T. D. (2007) Regulation of actin filament assembly by Arp2/3 complex and formins. *Annu. Rev. Biophys. Biomol. Struct. 36*, 451-477.

(25) Nolen, B. J., and Pollard, T. D. (2007) Insights into the influence of nucleotides on actin family proteins from seven structures of Arp2/3 complex. *Mol. Cell 26*, 449-457.

(26) Nolen, B. J., Littlefield, R. S., and Pollard, T. D. (2004) Crystal structures of actin-related protein 2/3 complex with bound ATP or ADP. *Proc. Natl. Acad. Sci. 101*, 15627-15632.

(27) Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science 289*, 905-20.

(28) Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell 102*, 615-23.

(29) Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H. F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell 126*, 1065-77.

(30) Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science 285*, 2095-104.

(31) Jenni, S., Leibundgut, M., Maier, T., and Ban, N. (2006) Architecture of a fungal fatty acid synthase at 5 Å resolution. *Science 311*, 1263-7.

(32) Maier, T., Jenni, S., and Ban, N. (2006) Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science 311*, 1258-62.

(33) Robert L. White, I. D. S. (1982) Thiamin biosynthesis in yeast. Origin of the five-carbon unit of the thiazole moiety *J. Am. Chem. Soc. 104*, 4934-43.

(34) White, R. L., and Spenser, I. D. (1979) Thiamin biosynthesis in *Saccharomyces cerevisiae*. Origin of carbon-2 of the thiazole moiety. *Biochem. J. 179*, 315-25.

(35) Robert L. White, I. D. S. (1979) Biosynthesis of vitamin B1 in yeast. Origin of the thiazole unit *J. Am. Chem. Soc. 101*, 5102-5104.

(36) Wightman R. and Meacock, P. A. (2003) The THI5 gene family of *Saccharomyces cerevisiae*: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine. *Microbiology 149*, 1447-1460.

(37) Haas, A. L., Laun, N. P., and Begley, T. P. (2005) Thi20, a remarkable enzyme from *Saccharomyces cerevisiae* with dual thiamin biosynthetic and degradation activities. *Bioorg. Chem. 33*, 338-44.

(38) Webb, M. E., Marquet, A., Mendel, R. R., Rebeille, F., and Smith, A. G. (2007) Elucidating biosynthetic pathways for vitamins and cofactors. *Nat. Prod. Rep. 24*, 988-1008.

(39) Nosaka, K., Nishimura, H., Kawasaki, Y., Tsujihara, T., and Iwashima, A. (1994) Isolation and characterization of the THI6 gene encoding a bifunctional thiamin-phosphate pyrophosphorylase/hydroxyethylthiazole kinase from *Saccharomyces cerevisiae*. *J. Biol. Chem. 269*, 30510-6.

(40) Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA 9*, 644-7.

(41) Machado, C. R., Praekelt, U. M., de Oliveira, R. C., Barbosa, A. C., Byrne, K. L., Meacock, P. A., and Menck, C. F. (1997) Dual role for the yeast THI4 gene in thiamine biosynthesis and DNA damage tolerance. *J. Mol. Biol. 273*, 114-21.

(42) Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2006) Thiamin biosynthesis in eukaryotes: characterization of the enzyme-bound product of thiazole synthase from *Saccharomyces cerevisiae* and its implications in thiazole biosynthesis. *J. Am. Chem. Soc. 128*, 7158-9.

(43) Jurgenson, C. T., Chatterjee, A., Begley, T. P., and Ealick, S. E. (2006) Structural insights into the function of the thiamin biosynthetic enzyme Thi4 from *Saccharomyces cerevisiae*. *Biochemistry 45*, 11061-70.

(44) Godoi, P. H., Galhardo, R. S., Luche, D. D., Van Sluys, M. A., Menck, C. F., and Oliva, G. (2006) Structure of the thiazole biosynthetic enzyme THI1 from *Arabidopsis thaliana*. *J. Biol. Chem. 281*, 30957-66.

(45) Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2007) Biosynthesis of thiamin thiazole in eukaryotes: conversion of NAD to an advanced intermediate. *J. Am. Chem. Soc. 129*, 2914-22.

(46) Ealick, S. E., and Begley, T. P. (2007) Biochemistry: molecular cannibalism. *Nature 446*, 387-8.

(47) Becker, M. A., Kredich, N. M., and Tomkins, G. M. (1969) Purification and characterization of O-acetylserine sulfhydrylase-A from *Salmonella typhimurium*. *J. Biol. Chem. 244*, 2418-27.

(48) Schnell, R., Oehlmann, W., Singh, M., and Schneider, G. (2007) Structural Insights into Catalysis and Inhibition of O-Acetylserine Sulfhydrylase from *Mycobacterium tuberculosis*: Crystal structures of the enzyme α-aminoacrylate intermediate and an enzyme-inhibitor complex. *J Biol. Chem. 282*, 23473-23481.

(49) Russell, D. G. (2003) Phagosomes, fatty acids and tuberculosis. *Nat. Cell. Biol. 5*, 776-8.

(50) Sibley, L. D., Adams, L. B., and Krahenbuhl, J. L. (1990) Inhibition of interferon-γ-mediated activation in mouse macrophages treated with lipoarabinomannan. *Clin. Exp. Immunol. 80*, 141-8.

(51) Chatterjee, D., Roberts, A. D., Lowell, K., Brennan, P. J., and Orme, I. M. (1992) Structural basis of capacity of lipoarabinomannan to induce secretion of tumor necrosis factor. *Infect. Immun. 60*, 1249-53.

(52) Ozeki, Y., Kaneda, K., Fujiwara, N., Morimoto, M., Oka, S., and Yano, I. (1997) *In vivo* induction of apoptosis in the thymus by administration of mycobacterial cord factor (trehalose 6,6´-dimycolate). *Infect. Immun. 65*, 1793-9.

(53) Vachula, M., Holzer, T. J., and Andersen, B. R. (1989) Suppression of monocyte oxidative response by phenolic glycolipid I of *Mycobacterium leprae*. *J. Immunol. 142*, 1696-701.

(54) Ng, V., Zanazzi, G., Timpl, R., Talts, J. F., Salzer, J. L., Brennan, P. J., and Rambukkana, A. (2000) Role of the cell wall phenolic glycolipid-1 in the peripheral nerve predilection of *Mycobacterium leprae*. *Cell 103*, 511-24.

(55) Li, Z., Kelley, C., Collins, F., Rouse, D., and Morris, S. (1998) Expression of katG in *Mycobacterium tuberculosis* is associated with its growth and persistence in mice and guinea pigs. *J. Infect. Dis. 177*, 1030-5.

(56) Manganelli, R., Voskuil, M. I., Schoolnik, G. K., Dubnau, E., Gomez, M., and Smith, I. (2002) Role of the extracytoplasmic-function σ Factor σH in *Mycobacterium tuberculosis* global gene expression. *Mol. Microbiol. 45*, 365-374.

(57) Burns, K. E., Baumgart, S., Dorrestein, P. C., Zhai, H., McLafferty, F. W., and Begley, T. P. (2005) Reconstitution of a new cysteine biosynthetic pathway in *Mycobacterium tuberculosis*. *J. Am. Chem. Soc. 127*, 11602-3.

(58) Lehmann, C., Begley, T. P., and Ealick, S. E. (2006) Structure of the *Escherichia coli* ThiS-ThiF complex, a key component of the sulfur transfer system in thiamin biosynthesis. *Biochemistry 45*, 11-9.

(59) Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V., and Schindelin, H. (2001) Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. *Nature 414*, 325-9.

(60) Erion, M. D., Stoeckler, J. D., Guida, W. C., Walter, R. L., and Ealick, S. E. (1997) Purine nucleoside phosphorylase. 2. Catalytic mechanism. *Biochemistry 36*, 11735-48.

(61) Mao, C., Cook, W. J., Zhou, M., Federov, A. A., Almo, S. C., and Ealick, S. E. (1998) Calf spleen purine nucleoside phosphorylase complexed with substrates and substrate analogues. *Biochemistry 37*, 7135-46.

(62) Huber, R., Langworthy, T. A., Koenig, H., Thomm, M., Woese, C. R., Sleytr, U. B., and Stetter, K. O. (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol. 144*, 324-33.

(63) Razvi, A., and Scholtz, J. M. (2006) Lessons in stability from thermophilic proteins. *Protein Sci. 15*, 1569-78.

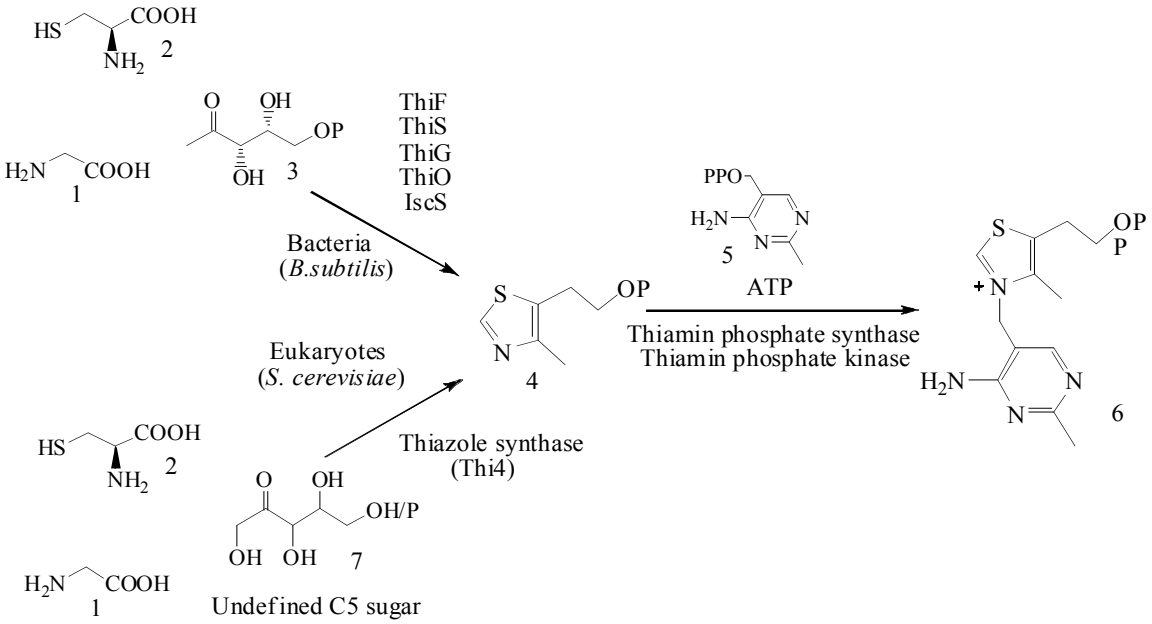
(64) Sadeghi, M., Naderi-Manesh, H., Zarrabi, M., and Ranjbar, B. (2006) Effective factors in thermostability of thermophilic proteins. *Biophys. Chem. 119*, 256-70.

CHAPTER TWO

STRUCTURAL INSIGHTS INTO THE FUNCTION OF THE THAMIN BIOSYNTHETIC ENZYME THI4 FROM SACCHAROMYCES CEREVISIAE*

Section 2.1 Introduction

Thiamin pyrophosphate **6** plays a key role in amino acid and carbohydrate metabolism and is an essential cofactor in all living systems. Its biosynthesis requires the separate formation of the pyrimidine **5** and thiazole **4** heterocycles, which are then coupled to form the cofactor (Scheme 1). The mechanism of the thiazole formation in bacteria has been elucidated and involves a complex oxidative condensation where five

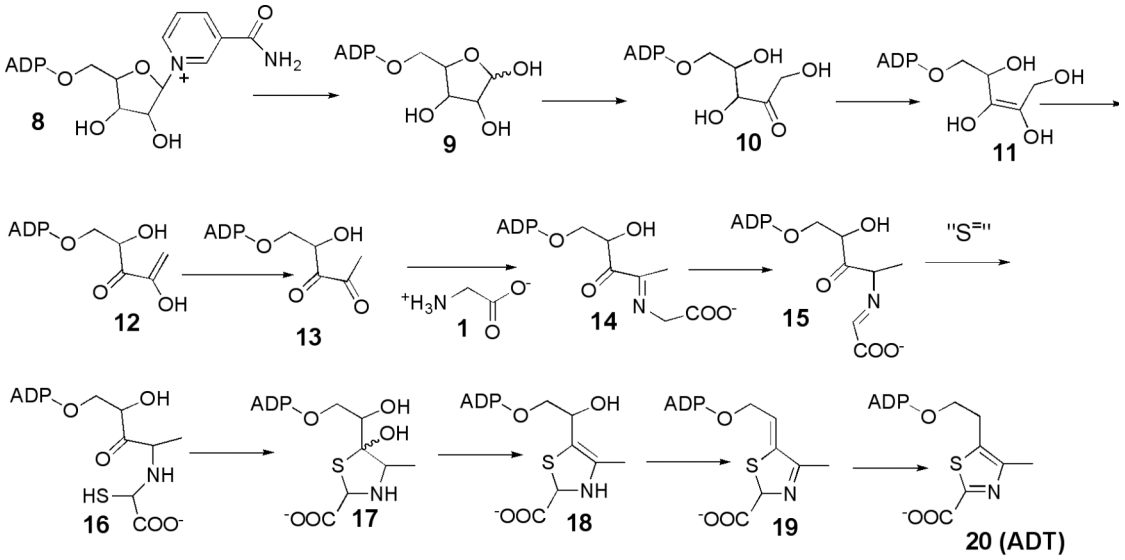


Scheme 2.1

different enzymes catalyze reactions involving three different substrates (1-deoxy-D-xylulose-5-phosphate **3**, cysteine **2** and glycine **1** or tyrosine) (*1*). The bacterial biosynthesis of the pyrimidine moiety is still poorly understood.

In contrast to the bacterial system, thiamin biosynthetic studies in eukaryotes are still at an early stage. Labeling studies in *Saccharomyces cerevisiae* have demonstrated that the thiamin thiazole is biosynthesized from an unidentified five carbon carbohydrate **7**, glycine **1** and cysteine **2** (*2-4*) (Scheme 1). Only one eukaryotic thiazole biosynthetic enzyme has been identified (Thi4) (*5*). This enzyme is a 326 amino acid (35 kDa) cytosolic protein. Sequence analysis suggests a conserved dinucleotide binding motif but does not reveal a biochemical function.

Our recently published results suggest an enzymatic mechanism whereby the cofactor nicotinamide adenine dinucleotide (NAD) is converted into adenosine diphospho-5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid (ADT) (*6*). The ADT ligand is tightly bound and can only be released *in vitro* upon protein denaturation. In



Scheme 2.2

this mechanism (Scheme 2), cleavage of the N-glycosyl bond of NAD **8** would give ADP ribose **9**. Ring opening of this hemi-acetal followed by tautomerization would give **10**. A second tautomerization followed by loss of water would give **12**. Tautomerization of **12** to the diketone **13** followed by imine formation with glycine **1** would give **14**. Imine tautomerization to **15** sets up the system to react with the sulfur donor, depicted here as sulfide to give **16**. Cyclization gives **17**. Two dehydration reactions followed by a final tautomerization complete the thiazole formation in ADT **20**. This proposal suggests that the first step in thiazole formation is similar to the chemistry used in ADP ribosylation.

Here, we report the structure of the *S. cerevisiae* thiazole synthase which reveals the nature of binding to ADT found in the enzyme active site. The identification of this product suggests that NAD is the source of the carbohydrate precursor and provides the first insight into the mechanism of thiamin thiazole biosynthesis in eukaryotes.

Section 2.2 Materials and Methods

Protein Production

The overexpression vector for the *THI4* gene was constructed at the Cornell Protein Facility by standard methods from genomic DNA harvested from cells obtained from the American Type Culture Collection. *THI4* was cloned into a pET28 overexpression vector encoding a 6x polyhistidine tag with a thrombin cleavage site located between the tag and the protein. The vectors were transformed into B834(DE3) (Novagen) methionine auxotrophic *Escherichia coli* cells (Novagen). Cells were grown in LB media at 37 °C until an OD_{600} of 1.0 was reached. The cells were induced using 1 mM isopropyl-β-D-thiogalactopyranoside and were incubated for 16-18 hours at 18 °C. Cell pellets that were collected after induction were placed into a beaker containing 50 mL of buffer with 5 mM imidazole, 0.5 M NaCl and 20 mM Tris at pH 7.9 (lysis

buffer) and lysed using sonication. The cell lysate was separated from the insoluble cell particles by centrifugation at 58,000 g at 4 °C for 20 min. The supernatant was decanted and loaded onto a Ni-NTA affinity column equilibrated with the same buffer used in the lysis step. The column with bound Thi4 was washed with lysis buffer containing 60 mM imidazole to remove any remaining proteins that may have affinity to the Ni column. The protein was eluted by washing the column with lysis buffer containing 0.5 M imidazole. The eluate was buffer exchanged into 20 mM Tris at pH 7.9 using Bio-Rad DG size exclusion chromatography columns with a molecular weight cut off of 6 kDa as per the manufacturer's instructions. The resulting solution was then concentrated in an Amicon Ultra centrifugal filter with a molecular weight cut off of 10 kDa by spinning at 5,000 g for 30-45 min at 4 °C until the concentration reached 10-20 mg/mL as measured using the Bradford Assay. The polyhistidine tag was removed using biotinylated thrombin from the Novagen Thrombin Cleavage Capture kit as per the manufacturer's instructions. Biotinylated thrombin was removed by adding streptavidin agarose beads to the reaction mixture, allowing it to equilibrate for 30 min, then collecting the beads through a spin filter. The cleaved histidine tags were separated by running the filtrate through a Ni-NTA column equilibrated with 60 mM imidazole, 0.5 M NaCl and 20 mM Tris at pH 7.9. The eluate was then buffer exchanged into 20 mM Tris at pH 7.9 and concentrated to 10-20 mg/mL as described in the buffer exchange step with histidine tagged Thi4. The final Thi4 protein product had the endogenously bound ligand ADT, which became apparent during structure determination and refinement.

Crystallization and X-ray Data Collection

The hanging drop vapor diffusion method was used to grow crystals of Thi4. The well solutions consisted of 1.65-2.00 M Li_2SO_4 and 0.1 M HEPES at pH 7.5. A hanging drop containing 1 μL of 10 mg/mL of the polyhistidine tag cleaved Thi4 and

2 μL of well solution were mixed and incubated at 22 °C. Cubic crystals appeared in two to four weeks and grew 100-200 μm along each edge. Crystals were mounted onto a cryoloop and cryprotected by dipping into a saturated solution of Li_2SO_4 as prescribed by Rubinson *et al. (7)* before being plunged into liquid N_2. Data were collected on the A1 beamline at the Cornell High Energy Synchrotron Source using a Quantum 210 CCD detector (Area Detector Systems Corporation). Individual frames were collected in 0.5 degree oscillations for 220 frames with an exposure time of 60 s per frame under a cryostream of liquid N_2 with the crystal mounted 190 mm from the detector. Integration and scaling of the data were done using the HKL2000 suite of programs (*8*). Data collection statistics are shown in Table 2.1.

Table 2.1. Data collection Statistics

Resolution (Å)	1.82
Space group	I4
Unit-cell parameters (Å)	
a	140.70
c	73.33
No. of reflections	280639
No. of unique reflections	63286
Redundancy	4.4 (3.1)
Completeness	98.5 (94.0)
R_{sym} [a] (%)	25.6 (4.5)
I/σ(I)	22.7 (3.3)

Numbers in parentheses are from the highest resolution shell.

[a] $R_{sym} = \Sigma\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma\langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices hkl.

Structure Determination

The structure of Thi4 was determined using the method of molecular replacement with the structure of Thi1 from *Arabidopsis thaliana* (PDB code 1RP0) as the molecular replacement search model. Refinement was carried out using rigid-body refinement followed by simulated annealing using CNS (*9*) as well as REFMAC

Table 2.2. Refinement Statistics

Resolution (Å)	1.82
No. of non-H atoms	5192
No. of protein atoms	4516
No. of water atoms	598
No. of ligand atoms	76
R factor (%) [a]	19.6
R_{free} (%) [b]	23.1
Ramachandran plot	
Most favored region (%)	90.7
Additionally allowed regions (%)	7.4
Generously allowed region (%)	1.0
Disallowed region (%)	1.0
r.m.s.deviations from ideal	
Bonds (Å)	0.027
Angles (°)	2.2

[a] R factor = Σ | $|F_{obs}| - K|F_{calc}|$ | / Σ_{hkl} $|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively

[b] R_{free} was calculated using 5% of all reflections that were excluded at all stages of refinement.

from CCP4 (*10*) and model rebuilding using the graphics programs O (*11*) and COOT (*12*). The final refinement statistics are shown in Table 2.2. All graphics were prepared using PyMOL (*13*) and ChemDraw (CambridgeSoft).

Section 2.3 Results

Monomer Structure

Thi4 is a homooctamer with two monomers, designated A and B, per asymmetric unit. The N-terminal regions of chains A and B are disordered from residue 1 to 16 and 1 to 15, respectively. Other areas that have disordered loops for chain A include residues 180-184, 200-202 and 239-240 and for chain B include residues 179-183, 199-202 and 238-241. The Thi4 monomer consists of ten β-strands and nine α-helices. The labeled monomer and a topology diagram representation are shown in Figure 2.1. The monomer consists of a central five-stranded parallel β-sheet flanked on one side by three α-helices, and on the other by an anti-parallel three-stranded β-sheet that lies between 8-9 Å above the plane of the central β-sheet, and is rotated by an angle of approximately 30 degrees. Long α-helices are found on both the N-terminal and C-terminal ends of the monomer. The N-terminal amphipathic helix spans residues 45-62. The C-terminal helix spans residues 307-324 and is also amphipathic. It is located in a cleft of the protein in close proximity to helices α7 and α2 and is involved in numerous interactions with the N-terminal region of the adjacent monomer. The N-terminal region is largely disordered with the first 16 residues of chain A and 15 residues of chain B missing from the electron density. The monomer chain contains two 3_{10}-helices with the first α-helix beginning at residue 45. Thirty-one of the first 44 residues are either hydrophilic or alanine, with the few large hydrophobic residues, such as Phe38 and Phe40, pointing toward the hydrophobic core of the protein. The largely hydrophilic nature of the N-terminus may be due to its location on the surface of the octomeric structure, thus lending to the solubility of the protein.

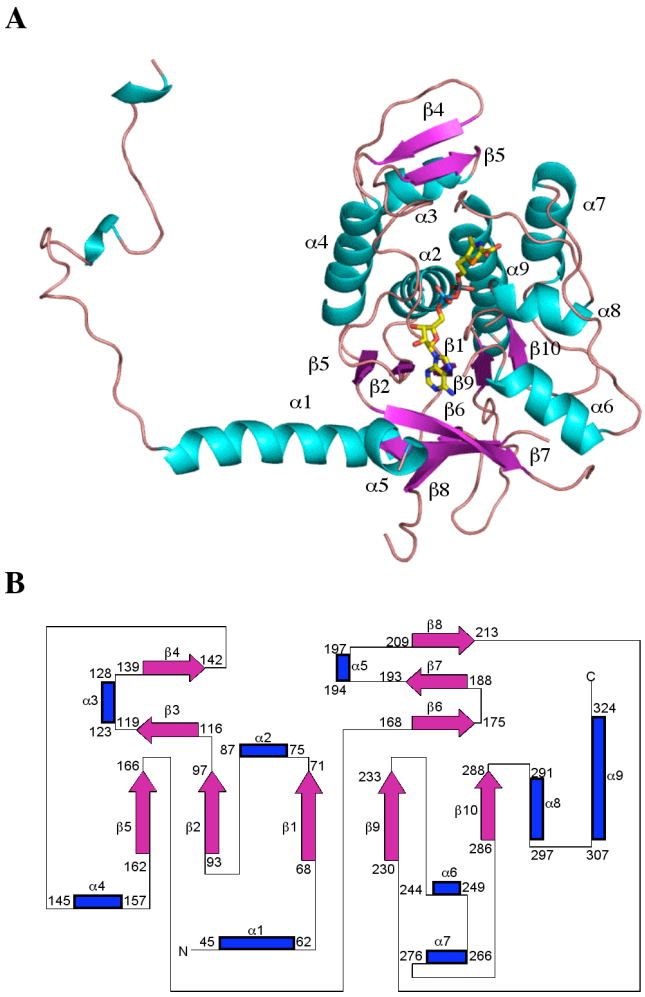


Figure 2.1. Structure of the Thi4 monomer. A) Ribbon diagram showing ten β-strands (magenta) and nine α-helices (blue). The core of the monomer consists of a parallel β-sheet made of strands 1, 2, 5, 9 and 10 which are sandwiched between an antiparallel β-sheet below made of strands 6, 7 and 8, and an α-helical bundle made of helices 2, 3, 4, 8, 7 and 9 from above. ADT is represented in stick form in the active site. B) Topology diagram of Thi4 highlighting the core five stranded β-sheet.

Pro121 is found in the *cis* conformation and follows a loop consisting of residues 98-115. Within this region is a triglycyl group which gives increased flexibility to the loop. Isomerization of this proline could result in the breakage of some of the hydrogen bond contacts which hold ADT in place such as those contributed by Ser98 and Gly105, thereby reducing the binding energy of ADT and allowing for its release and further catalysis.

Octomer Structure

Thi4 exists as a homooctomer with the disordered and largely hydrophilic N-terminal regions located on the exterior of the molecule (Figure 2.2). The octomer has the shape of a ring with flattened sides with the inner and outer diameters measuring approximately 30 Å and 96 Å, respectively, and a thickness of approximately 60-70 Å. The active site of Thi4 is found in the 30 Å diameter inner ring, which allows sufficient room for the entry of substrate molecules. The outer and inner surfaces are lined with hydrophilic residues with the overall magnitude of the electrostatic potential being positive both inside and outside the ring. The octomeric complex may be viewed as a tetramer of dimers with many interdimeric contacts at the interface between two monomers. The other major interface forms between adjacent dimers. A stereoview of the dimeric complex is shown in Figure 2.3. Two long α-helices, one from each chain, are seen flanking either side of a twofold axis, and are arranged in antiparallel fashion. These helices correspond to α4 (Figures 2.1B and 2.3) and facilitate the formation of the dimer largely through hydrophobic and van der Waals interactions, with the exception of residues Thr151 and Ser154 which form short hydrogen bonds to Ser154 (2.71 Å) and Thr151 (2.76 Å) of the adjacent monomer, respectively. The α1 helix passes orthogonally to the twofold axis at residue 58, and

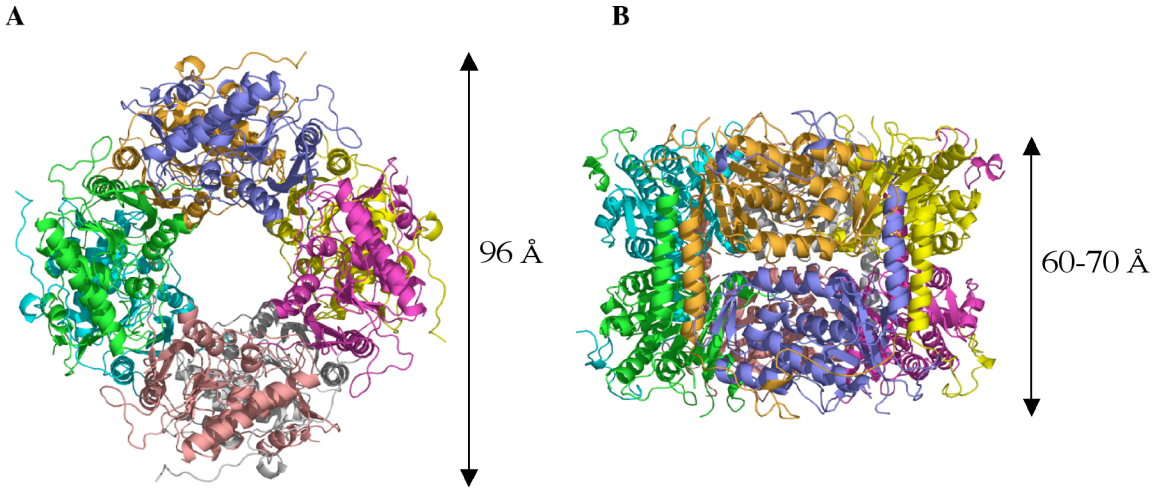


Figure 2.2. Structure of the Thi4 octomer. A) Ribbon diagram viewed through the central cavity. B) Ribbon diagram rotated 90° about the horizontal axis relative to panel (A). Each monomer is labeled in a different color.

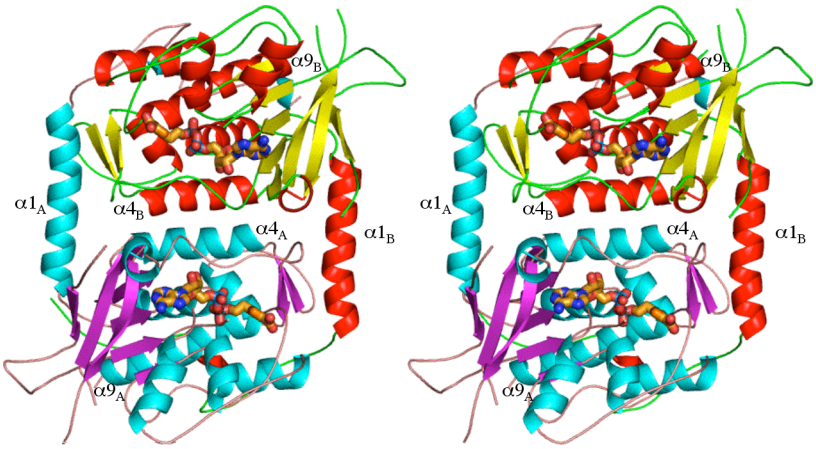


Figure 2.3. Stereoview of the tightly packed dimer. Helices and strands are colored in blue and magenta, respectively, in chain A; helices and strands are labeled in red and yellow, respectively, in chain B. The bound ligand is represented in stick form and can be seen in the active site of each monomer. The first and last secondary structural elements (α1 and α9) are labeled for each monomer. Helices α4 in each monomer are labeled to illustrate the main interaction in the dimer.

Table 2.3. Monomeric Interactions

A chain	B chain	Interaction
Tyr56, Phe57	Phe113	π-stacking
Ala41	Arg119	N-NH1 hydrogen bond
His24, Trp35	His315	π-stacking
Glu33	Lys322	Oε-Nζ hydrogen bond
His24	Glu318	Oε-Nζ hydrogen bond
Thr23	Asn87	N-Oδ hydrogen bond

interacts with the two short antiparallel β-strands β3 and β4 from chain B. The loop regions between β2 and β3 and α8 and α9 from chain B are also in close proximity to the N-terminal region. Table 2.3 lists hydrogen bonds and π-stacking interactions seen between chains A and B. The interface between individual dimers occurs mostly along the edge containing the α1 helix. The helix from one dimer (α1A) interacts with the α1 helix from the adjacent dimer (α1B) through three hydrogen bonding interactions between Arg50 and Asp61 of adjacent monomers. Additionally, the α1A

Table 2.4. Dimeric Interactions

A dimer	B dimer	Interaction
Arg44	Glu212	Oε-Nε hydrogen bond
Glu45	Arg176	Oε-NH_2 hydrogen bond
Glu45	Met266	Oε-NH hydrogen bond
Ser46	Glu212	Oδ-Oε hydrogen bond
Ser46	Arg50	O-N hydrogen bond
Ser49	Asn209	Oγ-Oδ hydrogen bond
Arg50	Asp61	Nε-Oδ hydrogen bond
Arg50	Asp61	Nε-O hydrogen bond
Arg50	Asp61	NH_2-O hydrogen bond

hydrogen bonds to β8, α7, and the loop region between β6 and β7 in the second dimer. A summary of these hydrogen bonding interactions is shown in Table 2.4. Most other interactions between dimers are the result of hydrophobic interactions.

Active site

The small molecule adenosine diphospho 5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid (ADT) (*6*) was bound to the Thi4 active site, which is located near the inner ring of the octomeric complex (Figure 2.4A). The adenosine moiety runs along the first β-strand in the fold characteristic of a dinucleotide binding site. The purine base is stabilized through π-stacking interactions between Phe241 and Phe244. Hydrogen bonding requirements for the N1 and N6 atoms of adenine are satisfied through the amide nitrogen atom and carbonyl oxygen atom of Val170, respectively. The N3 atom forms a hydrogen bond with the amide nitrogen atom of Ser98 at a distance of 3.22 Å. The 2′- and 3′-hydroxyl groups on the ribose ring are hydrogen bonded to Glu97, with the 2′-oxygen atom at a distance of 2.54 Å and 3′-oxygen atom at a distance of 2.82 Å from the nearest oxygen atom of the Glu97 side chain. The 2′-hydroxyl group also forms a hydrogen bond with Ser98 and a water molecule at a distance of 2.93 Å. The 3′-hydroxyl group forms additional hydrogen bonds with the amide nitrogen atom of Gly104 and another water molecule.

There are two hydrogen bonds to the O1 oxygen atom of the α-phosphate. One bond is to the amide nitrogen atom of Gly105, which is one of the three glycine residues seen in the triglycyl segment of the loop region connecting strands β2 and β3, and the other to a water molecule. Two hydrogen bonds with the α-phosphate O2 atom are formed through water molecules, but no bonds are made to side chains or the main chain. O1 from the β-phosphate forms three hydrogen bonds: one to a water molecule, and one each to the hydroxyl group and an amide nitrogen atom of Ser76. Two hydrogen bonds are formed by O2: one to a water molecule and the other to the

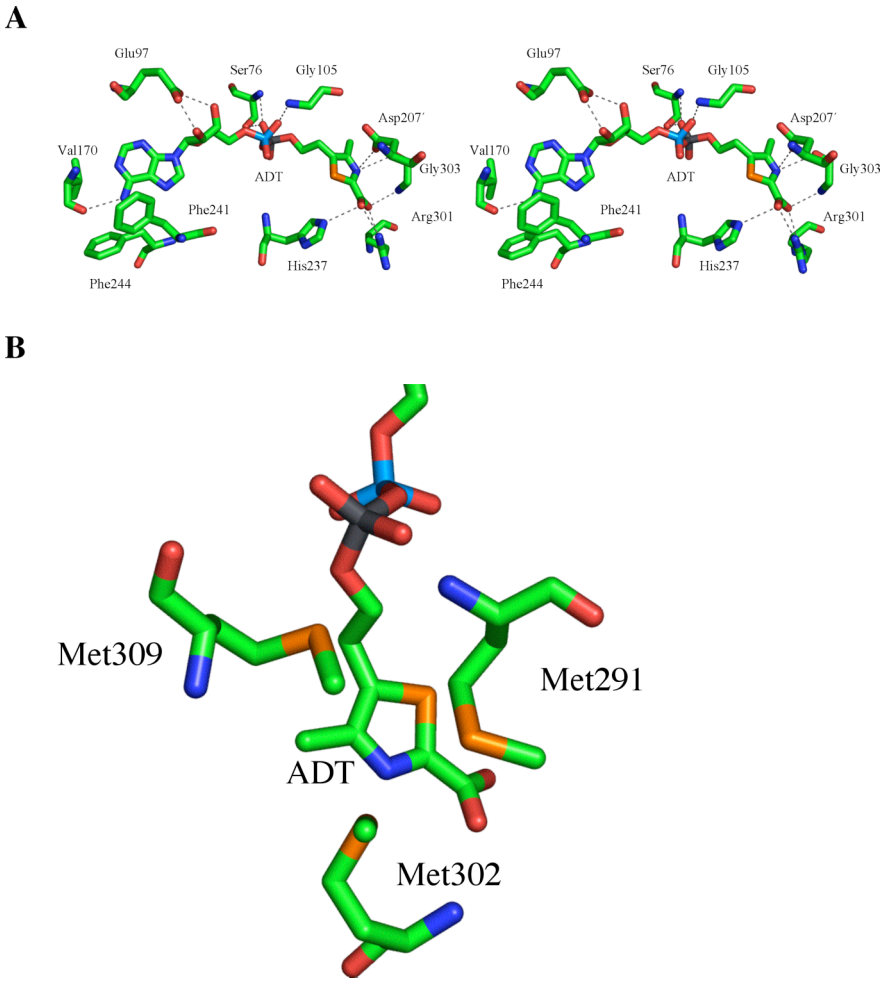


Figure 2.4 The Thi4 active site. A) Stereoview of the Thi4 ADT binding site. All amino acid side chains responsible for binding of ADT are shown. All residues are from a single chain, with the exception of Asp207′, which comes from another Thi4 monomer. Hydrogen bonds made between the residues and the ligand are represented as grey dashed lines. B) View of the thiazole moiety of ADT and the methionine triad which lies on one face of the heterocyclic ring.

amide nitrogen atom of Met291. The O3 atom which bridges the pyrophosphate group to the thiazole moiety is near the sulfur atom of Met309. This methionine together with Met291 and Met302 form an interesting three-fold symmetric triad just above the thiazole ring Figure 2.4B).

The thiazole moiety forms several hydrogen bonds with the protein, most of which are using the carboxylate group (Figure 2.4A). A single hydrogen bond exists between the nitrogen atom of the ring and a carbonyl group of Gly303. The carboxylate group forms two hydrogen bonds to the guanidinium moiety of Arg301, with each oxygen atom binding to a different nitrogen atom. The O1 atom from the carboxylate group also hydrogen bonds to the amide nitrogen atom of Gly303, while the O2 atom of the carboxylate group binds to Nε of His237. The thiazole ring is held in place through interactions with Asp207 from an adjacent dimer, which is positioned just above the ring facing the inner surface of the octomeric ring. One oxygen atom from Asp207 lies directly above the thiazole nitrogen at a distance of 3.29 Å while the other oxygen atom lies above the thiazole sulfur at a distance of 3.62 Å. The carboxylic acid of Asp207 forms a hydrogen bond with the carbonyl oxygen atom of Gly303 from another monomer in the octomeric complex. The other face of the thiazole ring is held in place by the methionine triad: Met291, Met302 and Met309 (Figure 2.4B).

The electron density maps for ADT show clear density for the ADP moiety, but are less clear in the region of the thiazole ring (Figure 2.5). The largest peak height corresponds to the sulfur atom as expected but the density is weaker than expected in the region of the N-C2 bond. The non-ideality of the final electron density map probably arises from the presence of ADT as well as minor contamination from one or two protein bound adenylated metabolites (*6*).

Section 2.4 Discussion

Comparison of Thi4 and the A. thaliana Thi1

The molecular search model used to determine the structure of the *S. cerevisiae* Thi4 was the ortholog of Thi4 in *A. thaliana* (PDB code 1RP0) Thi1 (*14*). The sequence identity between Thi1 and Thi4 is 67% with only 24 of 289 Thi1 residues in gap regions. Though the topology of Thi4 and Thi1 are the same, Thi4 has 326 residues per chain, while Thi1 has only 284. Two structural features account for most of this 42 amino acid difference. First, Thi4 contains an N-terminal extension of 32 residues compared to Thi1 and second, Thi4 contains a 15 amino acid insertion relative to Thi1 between β8 and β9. The middle of the loop contains a mostly hydrophilic segment (KNETRMK), faces a solvent channel in the crystal structure and is located away from the active site. The loop is absent in most Thi4 orthologs and probably does not directly affect the catalytic function.

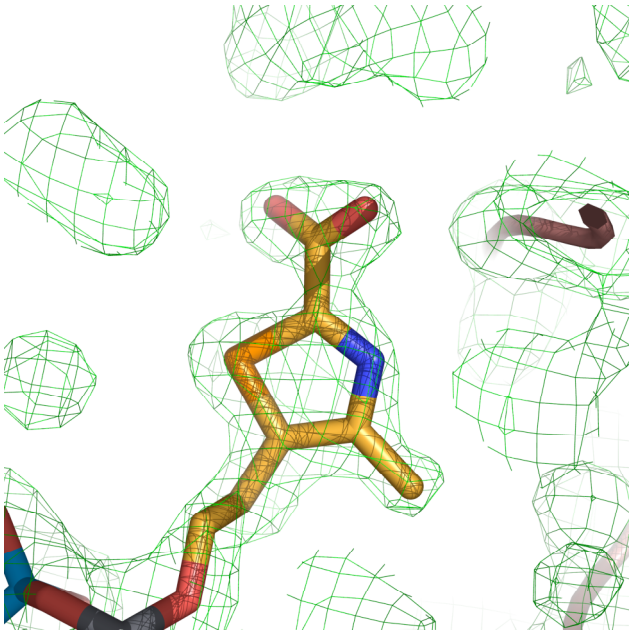


Figure 2.5. Electron density for ADT. The $2F_o$-F_c map is contoured at 1.0 σ. The figure was reproduced from Chatterjee, *et al*. (*6*).

Comparison of the Thi4 and Thi1 active sites reveals a tightly conserved arrangement around the thiazole moiety, despite a rotation of the thiazole ring of approximately 170° about the C5-C6 bond between the two structures. Figure 2.6 shows a superposition of ADT in both Thi1 and Thi4, highlighting residues that interact with the thiazole moiety. Neither of the sulfur atoms in either structure form hydrogen bonds or close contacts with the surrounding residues, but the hydrogen bond between the thiazole nitrogen atom and Gly303 in the Thi4 structure is absent in Thi1. Because the thiazole ring in Thi1 is rotated, the nitrogen atom in the thiazole of Thi1 binds to His189, Asp190, and a Zn^{2+} ion. Asp238, which corresponds to Asp190 in Thi1, has poor sidechain density in the A chain, and therefore has been modeled as alanine. Also, there is no density at all for Asp238 in the B chain of Thi4, so a comparison of binding between proteins for this residue is not possible. However, the

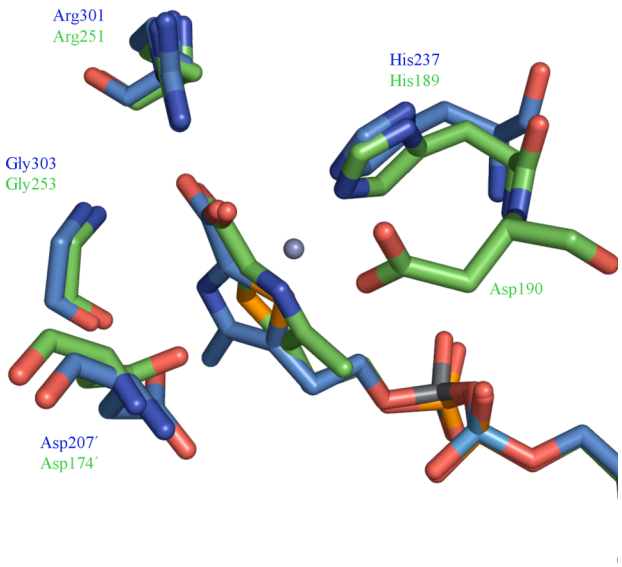


Figure 2.6. Overlay of Thi1 (green) and Thi4 (blue) active site residues in contact with the thiazole moiety. Asp207′/Asp174′ contributes from an adjacent Thi4/Thi1 monomer. The Zn^{2+} in Thi1 is shown as a blue sphere.

electron density peak for the ADT sulfur atom in the Thi4 complex is very high, so the conformation about the C5-C6 bond is well defined. The electron density for Thi1 shows no density for the C5 atom, which may indicate that different metabolites are present in the two structures.

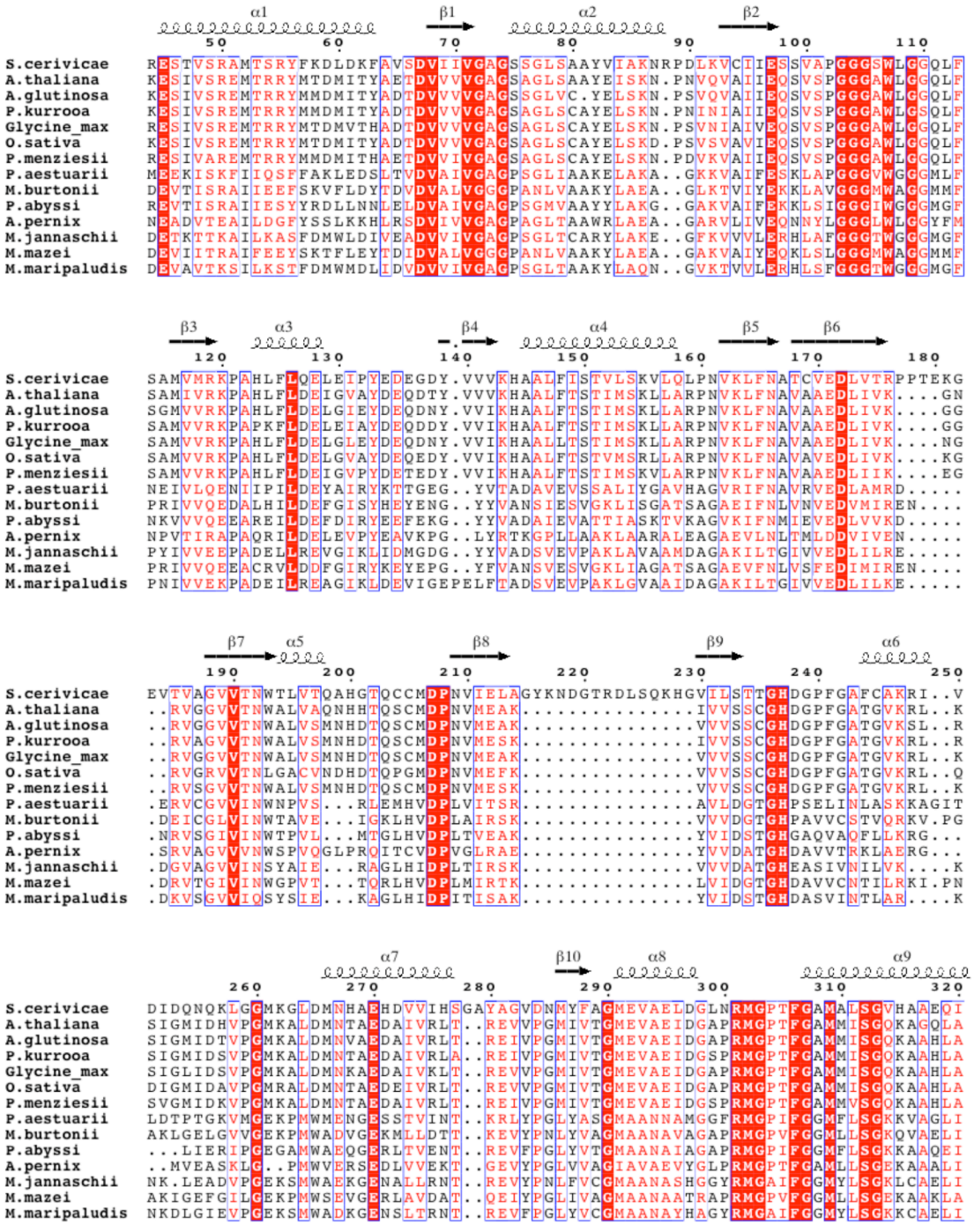
Conserved residues

A sequence alignment carried out with 14 different orthologs of Thi4 using ClustalW (*15*) along with secondary structural features is shown in Figure 2.7. Of those conserved residues, most are involved in interactions either within the protein or with its bound ligand. A conserved region in Thi4 corresponding to residues 68-78 aligns well with a dinucleotide binding motif (*16*). The characteristic motif of ϕϕGXGXXG, where ϕ is any hydrophobic residue and X is any residue, is retained among all the orthologs, with the exception of *Methanococcoides burtonii* which replaces the third glycine with an asparagine. Many entirely conserved residues are involved in ligand binding including Glu97, Gly105, Asp207, His237 and Arg301 in Thi4. Arg301 forms a hydrogen bond to both oxygen atoms on the carboxylate group attached to C2 of the thiazole ring. Gly105 is part of a triglycyl region within a flexible loop, with the amide nitrogen of Gly105 satisfying a hydrogen bond with an oxygen atom on the α-phosphate of ADT. Asp207 is found positioned above the thiazole ring with one oxygen atom at a distance of 3.29 Å while the other oxygen atom lies above the thiazole sulfur atom at a distance of 3.62 Å. The carboxylate of Asp207 is positioned by a water molecule, which bridges the carboxylate moiety and the carbonyl oxygen atom of Gly105 from an adjacent dimer of the octomeric complex. Gly303 is entirely conserved and hydrogen bonds with the thiazole ring through its backbone nitrogen atom. The structure reveals that Asp207 hydrogen bonds to Gly303 and positions its backbone nitrogen atom to hydrogen bond to the carboxylate group from the thiazole ring. His237 forms a hydrogen bond with the

carboxylate group of the thiazole moiety of ADT via the Nε atom, and may be involved in catalysis by acting as an active site base through activation by a water molecule which is hydrogen bonded to the Nδ nitrogen atom of the imidazole ring (see below). Glu270 forms hydrogen bonds with the hydroxyl group of Thr305 and the amide nitrogen of Ala308. Both of these residues are positioned after the loop containing Arg301, which is key in positioning the carboxylate group in ADT. Glu270 forms two hydrogen bonds to the guanidinium moiety of Arg119 which is the last residue in the last secondary structural element β3 that precedes the *cis*-proline Pro121.

Other exclusively conserved residues within the structure do not directly bind to the ADT molecule. Glu45 forms a salt bridge with Arg176 and makes a key dimer-dimer interaction to stabilize the octomeric complex. Though the arginine residue is not entirely conserved, it is usually replaced by lysine. Asp67 forms a hydrogen bond with the amide nitrogen of Lys92 and binds three water molecules which are on the outer surface of the octomer. This interaction serves to orient β2, which contains the Glu97 residue essential for positioning of the ribose ring of ADT. Gly109 forms a hydrogen bond with Phe113, but is also part of the flexible loop region, which also contains the triglycyl region. Asp172 plays a key role in positioning of Arg301 to bind the carboxylate group of ADT. First, Asp172 is hydrogen bonded to Arg248 which positions the side chain carboxylate in plane with the guanidinium moiety of Arg301. This in turn positions the guanidinium group to lie above the carboxylate of ADT with a geometry that favors strong hydrogen bonds. Though Arg248 is not entirely conserved, it is either arginine or lysine, either of which could orient Asp172 through a salt bridge. Trp193 from one monomer forms π-stacking interactions with a

Figure 2.7. Sequence alignment of Thi4 from *S. cerevisiae* with 14 Thi4 orthologs. The Thi4 secondary structures are listed above the corresponding sequence. Mostly conserved residues are colored pink, while those that are exclusively conserved are colored red.


α1 β1 α2 β2
S.cerivicae
A.thaliana
A.glutinosa
P.kurrooa
Glycine_max
O.sativa
P.menziesii
P.aestuarii
M.burtonii
P.abyssi
A.pernix
M.jannaschii
M.mazei
M.maripaludis
β3 α3 β4 α4 β5 β6
β7 α5 β8 β9 α6
α7 β10 α8 α9

Trp193 in another monomer and is buried amongst several aromatic residues, including Tyr56, Phe57, Phe113 and Phe165 from both chains. Conservation of Trp193 is probably key in octomer stabilization.

Mostly conserved Val170, Phe241, Phe244, Met291 and Gly303 interact with the ligand. Both phenylalanine residues form π-stacking interactions with the adenine moiety of ADT. Val170 contributes to hydrogen bonding interactions through its amide nitrogen and carbonyl oxygen atoms to N1 and N6 of the adenine ring, respectively. Though this residue is not entirely conserved, all Thi4 orthologs contain a small hydrophobic residue in this position suggesting that a hydrophobic environment near the side chains is important for orienting the hydrogen bonding main chain atoms. The methionine triad (Figure 2.4B) found in the helix-turn-helix motif encompassed by helices α8 and α9 is entirely conserved, with the exception of only one organism (*Aeropyrum pernix*) for which the equivalent of Met291 is isoleucine. None of the three methionine residues has significant side chain interactions with the exception of Met309, which forms a short contact of 3.3 Å between its sulfur atom and the bridging oxygen atom between the β-phosphate and the thiazole ring. The reason for the high conservation of this methionine triad is currently unknown.

Comparison with other protein structures

A structural similarity search using DALI (*17*) revealed that all of the top hits belong to the glutathione reductase type II family fold (GR_2) (*18*). A summary of Z scores and structural overlaps for the top 10 structures from the DALI output are shown in Table 2.5. The listed enzymes include phenol hydroxylase (PDB 1FOH), fumerate reductase (PDB, 1QJD) and the prokaryotic thiamin biosynthesis protein glycine oxidase (ThiO) (PDB, 1NG3), which have the same topology in the ADT binding region. This GR_2 family is characterized by a $GxGxxG(x)_nE/D$ sequence, where n can range from 16 to 19 residues in length. This sequence in Thi4

corresponds to residues 72 to 97, where n = 20, and the 97th residue is an exclusively conserved glutamic acid. The glutamic acid is responsible for hydrogen bonding to the 2′- and 3′- hydroxyl groups of the ribose sugar; a common feature in nucleotide

Table 2.5. DALI search results

PDB Code	Z-Score	LALI[a]	r.m.s.d.[b]	%ID[c]
1RP0	36.8	255	1.0	57%
1QJD	17.5	212	3.5	18%
1TRB	16.1	161	2.4	13%
1HYU	16.1	181	8.4	14%
1PBE	15.5	181	2.7	13%
1KDG	15.4	193	2.6	21%
1NG3	15.2	180	2.4	15%
1QLA	15.1	198	2.8	20%
1FOH	14.6	190	2.9	17%
1B8S	14.3	196	2.9	13%

[a] Total number of equivalent residues.

[b] Positional root mean square deviation of superimposed Cα in Angstroms.

[c] Percentage of sequence identity over equivalent positions.

binding proteins. The most unexpected result of Thi4 being a member of the GR_2 family is that every member to which Thi4 has a significant topological similarity via DALI is a flavoenzyme. However, our recently published studies suggest that Thi4 utilizes NAD as its substrate (*6*). Classic NAD-dependent oxidoreductases such as alcohol dehydrogenase, lactate dehydrogenase and 3-glyceraldehyde-3-phosphate dehydrogenase exhibit a Rossmann fold characteristic of NAD binding proteins (*19*), but the fold in Thi4 has a clearly different topology (Figure 2.1B).

A structural overlay between Thi4 and ThiO is shown in Figure 2.8. The FAD binding domain between Thi4 and ThiO overlaps well, with the rest of ThiO having neither any similarity to the dimeric interface between Thi4 monomers or to another Thi4 monomer within the octomeric complex. The ADT in Thi4 and FAD in ThiO overlap well in the ADP moiety, but diverge after the β-phosphate. Thus the GR_2 fold serves both as a cofactor binding topology for flavoenzymes and as a fold that can utilize NAD as a substrate to generate a critical component in yeast thiamin biosynthesis. When considering Thi4 from an evolutionary standpoint, it represents an unprecedented scenario where a protein fold that once served to bind FAD as a cofactor evolved to bind and catalyze NAD as a substrate without any topological rearrangement.

Mechanistic Implication

Previous studies, using labeled precursors, have demonstrated that the thiazole heterocycle of thiamin is biosynthesized in yeast from glycine, cysteine and an unidentified C5 sugar (*2-4*). The Thi4 structure with ADT bound at the active site shows that the unidentified C5 sugar is linked to ADP. This observation suggests that the C5 sugar precursor to the thiazole is derived from an adenylated metabolite synthesized in both *S. cerevisiae* and in *E. coli*, such as NAD, NADH, FAD or FADH

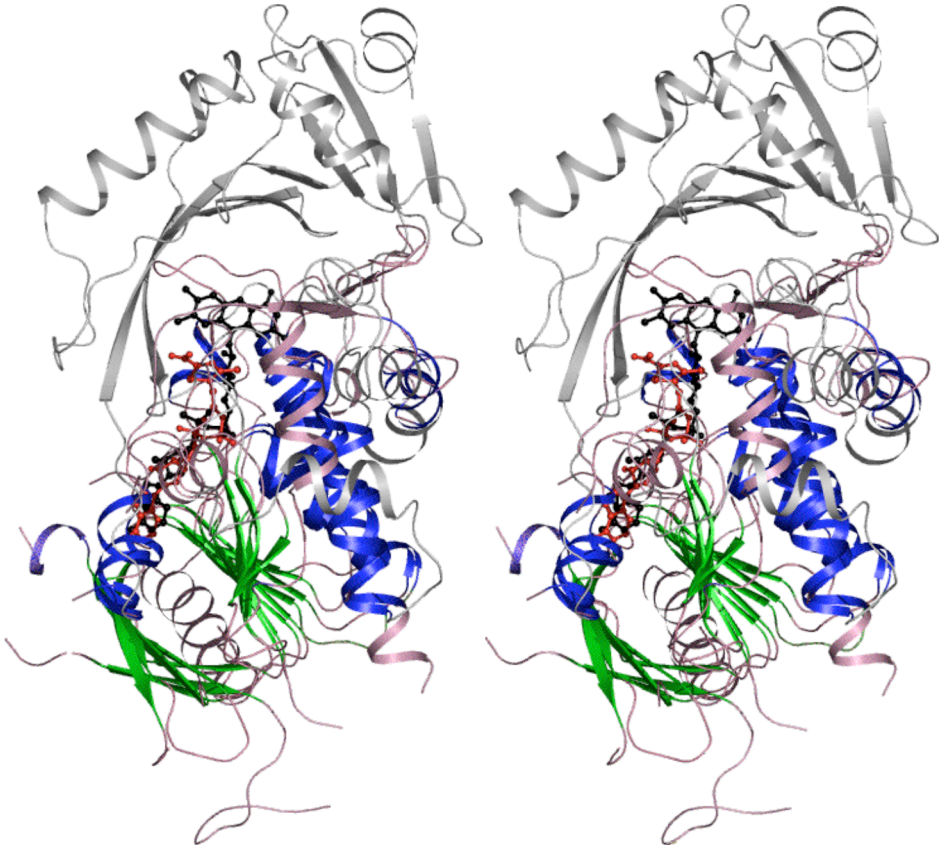


Figure 2.8. Stereoview of a superposition of Thi4 and ThiO showing the overlapping dinucleotide binding domains. Overlapping secondary structural elements are shown in blue (α-helices) or green (β-strands). Non-overlapping secondary structural elements are shown in pink (Thi4) or gray (ThiO). FAD bound to ThiO is shown as a black ball-and-stick model. ADT bound to Thi4 is shown as a red ball-and-stick model.

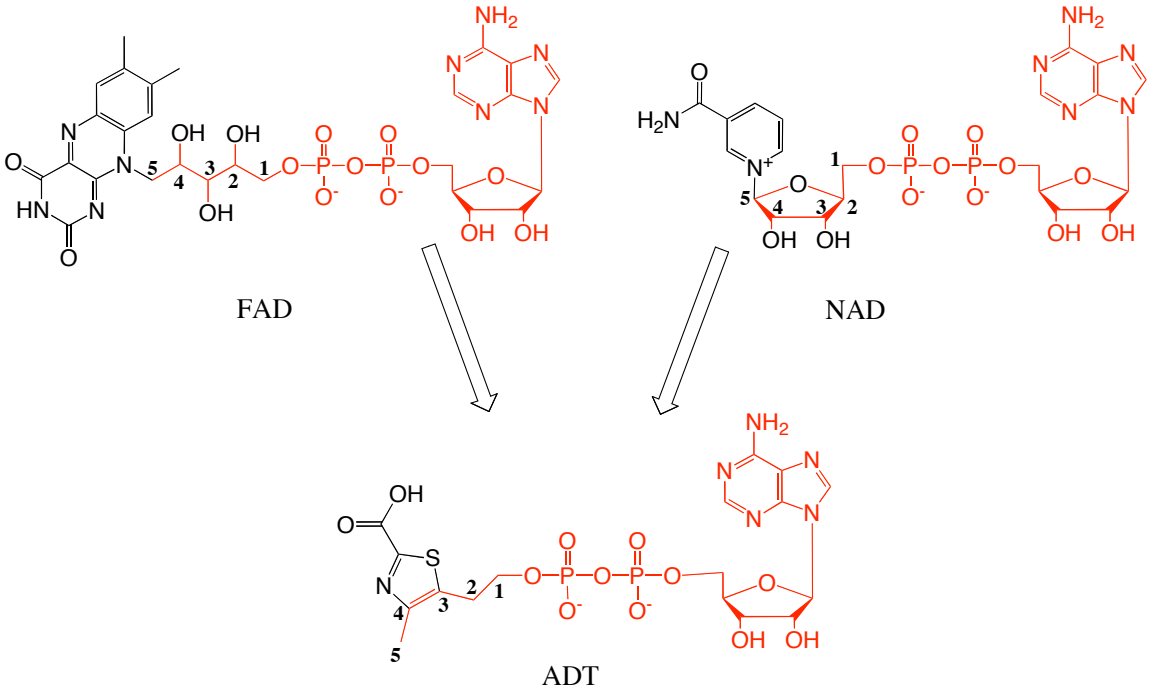


Figure 2.9. Chemical structures of FAD, NAD and ADT. Atoms common to all three structures are shown in red and the five carbon atoms corresponding to the thiazole are numbered. The remaining atoms of the thiazole are derived from glycine and an unidentified sulfur donor.

(Figure 2.9). NAD/NADH seems more likely than FAD/FADH because pyridine can be more readily cleaved from ribose than isoalloxazine from ribulose using an NAD-glycohydrolase-like activity. In addition, the pentose sugar of NAD is in the correct oxidation state to form the thiazole moiety, whereas in FAD/FADH it needs to be oxidized.

While the structure of the yeast thiazole synthase identifies NAD as the most likely thiazole precursor and provides key insights into the mechanism of thiazole formation in eukaryotes, much still remains to be discovered about this interesting protein. The mechanistic proposal outlined in Scheme 2 needs to be tested and refined and the sulfur source and the mechanism of sulfur transfer remain to be determined. Also the biological significance of the stable enzyme-product complex is unknown. In addition, four other properties of Thi4 and its orthologs are worth mentioning. In the exponential growth phase, Thi4 is an abundant protein (1.5% of the entire proteome) (*20*) in *N. crassa*, grown in the absence of thiamin. Genetic studies indicate that the *Arabidopsis* Thi4 (Thi1) protects mitochondrial DNA from damage (*21*). The *N. crassa* Thi4 (CyPBP37) is posttranslationally modified and forms a stable complex with a cyclophilin (NcCyP41) (*22*). The *Fusarium oxysporum* Thi4 (sti35) is induced by various stress conditions (*23*). These observations suggest that in addition to thiamin biosynthesis, Thi4 is likely to have additional functions in the cell

Section 2.5 Conclusions

Thiamin was the first vitamin discovered, but the last to have its biosynthesis elucidated, due to the novelty and complexity of its biosynthetic chemistry. While the major features of the bacterial pathway have now been elucidated, little is known about the biosynthesis of this important vitamin in eukaryotes. This paper describes the structure of the *S. cerevisiae* thiazole synthase (Thi4) with adenosine diphospho-5-(β-ethyl)-4-methyl-thiazole-2-carboxylic acid (ADT) bound at its active site. The

isolation of this product identifies NAD as the most likely precursor to the thiamin thiazole, suggests that its biosynthesis is mechanistically related to protein ADP ribosylation, which is an important cellular regulatory process, and provides the first mechanistic insights into the biosynthesis of the thiamin thiazole in eukaryotes. Additionally, the Thi4 structure reveals the first protein structure with a GR_2 domain that binds NAD instead of FAD, raising interesting questions about how this protein evolved from a flavoenzyme to an NAD binding enzyme.

BIBLIOGRAPHY

1. Settembre, E., Begley, T. P., and Ealick, S. E. (2003) Structural biology of enzymes of the thiamin biosynthesis pathway, *Curr. Opin. Struct. Biol. 13*, 739-47.
2. White, R. L., and Spenser, I. D. (1979) Thiamin biosynthesis in *Saccharomyces cerevisiae*. Origin of carbon-2 of the thiazole moiety, *Biochemical J. 179*, 315-25.
3. White, R. L., and Spenser, I. D. (1979) Biosynthesis of vitamin B1 in yeast. Origin of the thiazole unit, *J. Am. Chem. Soc. 101*, 5102-4.
4. White, R. L., and Spenser, I. D. (1982) Thiamin biosynthesis in yeast. Origin of the five-carbon unit of the thiazole moiety, *J. Am. Chem. Soc. 104*, 4934-43.
5. Praekelt, U. M., Byrne, K. L., and Meacock, P. A. (1994) Regulation of THI4 (MOL1), a thiamine-biosynthetic gene of *Saccharomyces cerevisiae*, *Yeast 10*, 481-90.
6. Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2006) Thiamin Biosynthesis in Eukaryotes: Characterization of the Enzyme-Bound Product of Thiazole Synthase from *Saccharomyces cerevisiae* and Its Implications in Thiazole Biosynthesis, *J. Am. Chem. Soc. 128*, 7158-9.
7. Rubinson, K. A., Ladner, J. E., Tordova, M., and Gilliland, G. L. (2000) Cryosalts: suppression of ice formation in macromolecular crystallography, *Acta Crystallogr. D 56*, 996-1001.
8. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.
9. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read,

R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D 54*, 905-21.

10. Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography, *Acta. Crystallogr. D 50*, 760-763.
11. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A 47 (Pt 2)*, 110-9.
12. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallog. D 60*, 2126-32.
13. DeLano, W. L. (2002), DeLano Scientific, San Carlos, CA.
14. Ribeiro, A., Praekelt, U., Akkermans, A. D., Meacock, P. A., van Kammen, A., Bisseling, T., and Pawlowski, K. (1996) Identification of agthi1, whose product is involved in biosynthesis of the thiamine precursor thiazole, in actinorhizal nodules of *Alnus glutinosa*, *Plant J. 10*, 361-8.
15. Pearson, W. R., and Lipman, D. J. (1988) Improved tools for biological sequence comparison, *Proc. Natl. Acad. Sci. U. S. A. 85*, 2444-8.
16. Moller, W., and Amons, R. (1985) Phosphate-binding sequences in nucleotide-binding proteins, *FEBS Lett. 186*, 1-7.
17. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123-38.
18. Dym, O., and Eisenberg, D. (2001) Sequence-structure analysis of FAD-containing proteins, *Protein Sci. 10*, 1712-28.
19. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein, *Nature 250*, 194-9.

20. Faou, P., and Tropschug, M. (2004) *Neurospora crassa* CyPBP37: a cytosolic stress protein that is able to replace yeast Thi4p function in the synthesis of vitamin B1, *J. Mol. Biol. 344*, 1147-57.

21. Machado, C. R., Praekelt, U. M., de Oliveira, R. C., Barbosa, A. C., Byrne, K. L., Meacock, P. A., and Menck, C. F. (1997) Dual role for the yeast THI4 gene in thiamine biosynthesis and DNA damage tolerance, *J. Mol. Biol. 273*, 114-21.

22. Faou, P., and Tropschug, M. (2003) A novel binding protein for a member of CyP40-type Cyclophilins: *N. crassa* CyPBP37, a growth and thiamine regulated protein homolog to yeast Thi4p, *J. Mol. Biol. 333*, 831-44.

23. Thanonkeo, P., Akiyama, K., Jain, S., and Takata, R. (2000) Targeted disruption of sti35, a stress-responsive gene in phytopathogenic fungus *Fusarium oxysporum*, *Curr. Microbiol. 41*, 284-9.

CHAPTER THREE

CRYSTAL STRUCTURE OF A SULFUR CARRIER PROTEIN COMPLEX FOUND IN THE CYSTEINE BIOSYNTHETIC PATHWAY OF MYCOBACTERIUM TUBERCULOSIS

Section 3.1 Introduction

CysM (Rv1336) and CysO (Rv1335) are two proteins that participate in a recently discovered cysteine biosynthesis pathway in *Mycobacterium tuberculosis* (*1*). In this pathway CysO thiocarboxylate displaces the acetyl group of O-acetylserine in a reaction catalyzed by the pyridoxal 5′-phosphate (PLP) dependent CysM protein. An N-S acyl shift, followed by Mec^{+} (Rv1334) catalyzed hydrolysis generates cysteine (Figure 3.1). This pathway stands in contrast to the widely used sulfide-dependent cysteine biosynthesis pathway also present in *M. tuberculosis* (CysK) and may be of importance in the oxidizing environment of the macrophage (*2-6*) because thiocarboxylates are much more resistant to oxidation than sulfide.

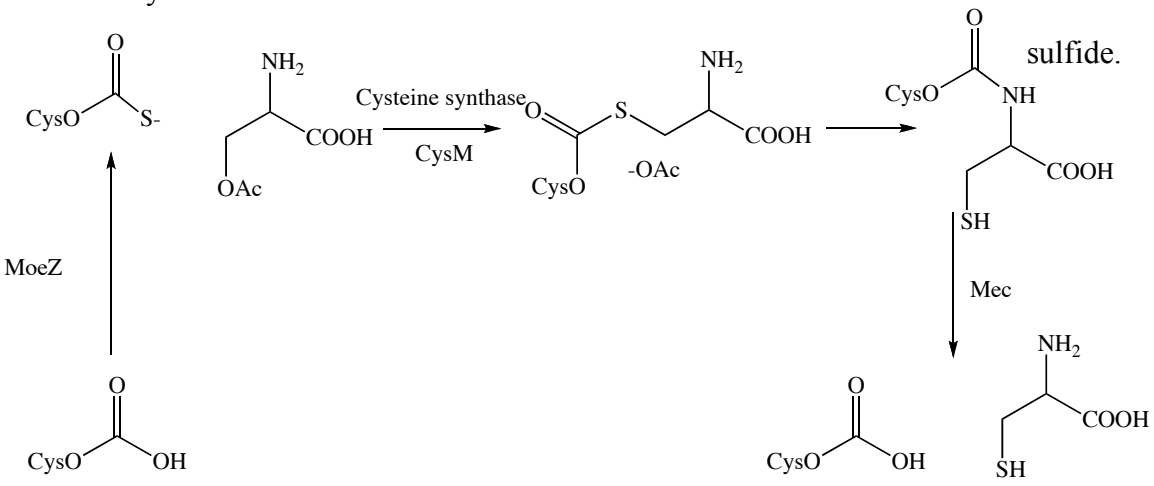


Figure 3.1. The new cysteine biosynthetic pathway in *M. tuberculosis*.

Consistent with this, the genes for CysM and CysO are upregulated when the organism is exposed to oxidative stress (*7*).

Sulfur-carrier proteins are structurally homologous to ubiquitin (*8*) and have now been identified in the biosynthetic pathways for cysteine (*1*), thiamin (*9*), molybdopterin (*10*), and thioquinolobactin (*11*). Like ubiquitin, the sulfur carrier protein is adenylated at a diglycyl C-terminus by a specific activating protein. The adenylated C-terminus is subsequently converted to a thiocarboxylate, which serves as the sulfide source. Members of the sulfur carrier protein family have diverse binding partners and show little sequence similarity. ThiS is the sulfur carrier protein in thiamin biosynthesis and is activated by ThiF (*12*). Thiocarboxylated-ThiS, deoxy-D-xylulose 5-phosphate and glycine imine serve as substrates for ThiG in the synthesis of the thiazole moiety of thiamin (*13*). The sulfur carrier protein MoaD is found in the molybdopterin biosynthetic pathway and its activating protein is MoeB (*14*). Crystal structures for the ThiF-ThiS (*15*), ThiG-ThiS (*16*), MoeB-MoaD (*17*) and MoaE-MoaD (*10*) complexes have been reported. An interesting variation of a sulfur carrier protein is found in the biosynthetic pathway of thioquinolobactin (*18*). In this system QbsE is the sulfur carrier protein and QbsC is the adenylating protein. QbsE has a diglycyl sequence followed by cysteine and phenylalanine at its C-terminus. Clustered with the genes for these two proteins is a gene for QbsD, a metal-dependent hydrolase, which cleaves the final two amino acids from QbsE generating the diglycyl C-terminus found in all of the other sulfur-carrier proteins.

We report here the X-ray crystal structure of the CysM-CysO complex of the cysteine biosynthetic pathway refined to 1.53 Å resolution, the structure of CysM alone at 2.8 Å resolution, and the structure of CysO alone at 2.7 Å resolution. The CysM-CysO complex is asymmetric with one molecule of CysO binding to one end of a CysM dimer. Conformational changes in CysM that occur upon complex formation

are described. The CysM-CysO interface is compared to the binding interfaces for ThiF-ThiS, ThiG-ThiS, MoeB-MoaD, MoaE-MoaD and ubiquitin in complex with the E1-like protein MMS2.

Section 3.2 Materials and Methods

Gene Cloning, Overexpression and Purification

M. tuberculosis DNA was a gift from Clifton Barry at the National Institutes of Health. pET-16b and pET-28a plasmids were purchased from Novagen. All plasmid DNA was purified with a DNA miniprep kit from Promega. A Perkin Elmer GeneAmp PCR system 2400 and Platinum Pfx DNA polymerase (Gibco Life Technologies) were used for PCR. Primers were purchased from Inegrated DNA Technologies. Oligonucleotides were prepared using *M. tuberculosis* DNA as a template for PCR, and sequencing was performed by the Bioresource Center at Cornell University. DNA fragments were purified by agarose gel electrophoresis followed by a QIAquick gel extraction kit from Qiagen. T4 DNA ligase was purchased from New England Biolabs. *Escherichia coli* strain DH5α was used as a recipient for transformation, propagation, and storage. Primers for the CysM plasmid were engineered to introduce NdeI and BamHI restriction sites at the 5′ and 3′ ends, respectively. The forward primer was 5′-GGG TGA GCG GAG CAT ATG ACA CGA TAC GAC-3′. The reverse primer was 5′-TTC GGA TCC GGC GGA TCC TCA TGC CCA TAG-3′. Primers for the CysO plasmid were engineered to introduce NdeI and XhoI restriction sites at the 5′ and 3′ ends, respectively. The forward primer was 5′-CCG AGA AAG GCC CAT ATG AAC GTC ACC GTA-3′. The reverse primer was 5′-TCG TGT CAT GTG CTC GAG TCA CCC ACC GGC-3′.

The CysM construct was transformed into B834(DE3) methionine auxotrophic *E. coli* (Novagen) for overexpression. Minimal media used for overexpression of B834(DE3) cells was prepared by dissolving 11.28 g of M9 salts along with 40 mg/L

of each amino acid excluding methionine, which was replaced by selenomethionine at 50 mg/L. The final media also contained 0.4% (w/v%) glucose, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 25 mg/mL $FeSO_4$, 100 mg/L ampicillin and 10 mL of 100x MEM vitamin solution (Invitrogen). Starter cultures were prepared by growing cells in 20 mL of LB media, then spinning the cells at 4800g for 5 min. The LB supernatant was decanted and the cell pellet resuspended with 20 mL of minimal media to inoculate 1 L of minimal media. Cells were grown at 37 °C until an OD600 of 1.0 was reached, followed by induction using 1 mM isopropyl-β-D-thiogalactopyranoside for 16-18 h at 18 °C. Cell pellets were collected by centrifugation at 6400g for seven minutes, and placed into a beaker containing 50 mL of a lysis buffer composed of 5 mM imidazole, 0.5 M NaCl and 20 mM Tris at pH 7.9 to be lysed using sonication. Cell lysate was separated from the insoluble cell particles by centrifugation at 58,000g at 4 °C for 20 min. The supernatant was run over a Ni-NTA affinity column equilibrated with lysis buffer. The column was washed with lysis buffer containing 60 mM imidazole and eluted with lysis buffer containing 1M imidazole. The eluate was buffer exchanged into 20 mM pH 7.9 Tris and 5 mM 1,4-dithiothreitol (DTT) using size exclusion chromatography columns (Bio-Rad DG) with a molecular weight cut off of 6 kDa. A centrifugal filter (Amicon Ultra) with a molecular weight cut off of 10 kDa was used to concentrate the solution by spinning at 5,000g at 4 °C. The final protein concentration was 10-20 mg/mL as measured using the Bradford Assay (*19*). SDS-PAGE analysis showed a purity of greater than 95%. Native CysM and CysO were transformed into BL21(DE3) cells, overexpressed in LB media and purified as described above, without the use of DTT.

In an attempt to improve crystallization, a putative surface residue in CysM (Lys204) was mutated to alanine using site-directed mutagenesis (*20*). A standard PCR protocol using *Pfu*Turbo DNA polymerase per the manufacturer's instructions

(Invitrogen) and *Dpn*I (New England Biolabs) to digest the methylated parental DNA prior to transformation was used. The forward primer was 5´- GCA CGT TGC CAA CGT CGC GAT CGT GGC GGC CGA ACC CCG C-3´ and the reverse primer was 5´- GCG GGG TTC GGC CGC CAC GAT CGC GAC GTT GGC AAC GTG C-3´. Clones were screened by restriction digest for the introduction of a *Pvu*I site. A representative clone with the correct restriction pattern was sequenced by the Bioresource Center at Cornell University.

Protein Crystallization and Cryoprotection

A solution of native CysM-CysO was prepared by mixing separate 15 mg/mL solutions each of CysO and CysM in a 2:1 volumetric ratio. This solution was screened using the hanging drop vapor diffusion method. Individual crystals grew in a range of 7-10% PEG 4,000 (wt/v%), 0.1 M sodium citrate pH 5.8 and 0.2 M ammonium acetate at 22 °C. Selenomethionine crystals grew in the same conditions with 5 mM DTT. Crystals were green in color and grew to 150 x 150 x 50 μm^3 within two to four weeks. The K204 mutant was also screened using the hanging drop method with a 10 mg/mL protein solution. Crystals of the K204A mutant grew to approximately 100 x 40 x 20 μm^3 in two to four weeks using the hanging drop method in 0.1 M HEPES pH 7.1 and 0.64 M Li_2SO_4 at 22 °C. Crystals of CysO were grown using the hanging drop method in the presence of the activating protein MoeZ (Rv3206). A solution of native MoeZ-CysO was prepared by mixing separate 10 mg/mL solutions each of MoeZ and CysO in a 1:1 volumetric ratio. Crystals appeared in 0.2 M $NaNO_3$ and 20% (wt/v%) PEG 3350 after approximately six months as colorless thick plates with dimensions of 200 x 40 x 40 μm^3. Preliminary analysis of the X-ray diffraction data showed that these crystals contained only CysO.

Cryoprotection of the CysM-CysO complex and CysM K204A mutant crystals was carried out through sequential dipping into well solutions with increasing

concentrations of ethylene glycol. The cryosolution for the selenomethionine derivative and native CysM-CysO crystals contained 9% PEG 4000, 0.1 M sodium citrate pH 5.8, 0.2 M ammonium acetate, 1 mM DTT and either 5% or 25% ethylene glycol by volume. Crystals were first looped out of the mother liquor, and then placed in the 5% ethylene glycol solution for approximately 10 s, followed by dipping into the 25% ethylene glycol solution for another 10 s. The crystal was then looped out of the 25% solution and plunged into liquid nitrogen for storage. Cryoprotection of the K204A CysM mutant was also done using sequentially increasing concentrations of ethylene glycol. CysM crystals were first dipped into a solution with 5% ethylene glycol, 0.64 M Li_2SO_4, and 0.1 M HEPES pH 7.1 followed by dipping into the same cryosolution with 25% ethylene glycol for approximately 10 s each. Crystals of CysO were cryoprotected by dipping into a solution of 0.2 $NaNO_3$, 20% PEG 3350 and 25% PEG 400 for approximately 10 s before plunging into liquid nitrogen.

X-ray Data Collection

Data for the selenomethionine derivative CysM-CysO crystal were collected at the 24-ID-C beamline at the Advanced Photon Source to 2.1 Å resolution. The energy was calibrated with Se foil, and an X-ray fluorescence scan was done on the CysM-CysO crystal to determine the absorption edge. Images were taken using a Quantum ADSC Q315 detector (Area Detectors Systems Corp.) for 720 frames with 1° oscillations in 10° wedges. Native data for the CysM K204A and CysM-CysO complex crystals were collected at the 8-BM beamline at APS using 1° oscillations for 140° and 300°, respectively, using a Quantum ADSC Q315 detector. Data for the CysO crystals were collected using a Rigaku RTP 300 RC rotating copper anode generator operating at 50 kV and 100 mA. A Rigaku RAXIS IV^{++} image plate detector was used for image detection. Data collection was done in 1° oscillations for a total of 215°. Data collection statistics for all datasets are summarized in Table 3.1.

Table 3.1. Data Collection Statistics

Sample	SeMet CysOM	Native CysOM[1]	K204A CysM	Native CysO
Source	APS 24ID	APS 8BM	APS 8BM	Home
Wavelength (Å)	0.97950	0.97949	0.94980	1.54178
Resolution (Å)	2.10	1.53	2.80	2.70
Space group	$P2_1$	$P2_1$	$P2_12_12_1$	C2
Unit-cell parameters				
a (Å)	56.2	55.8	72.4	87.9
b (Å)	80.6	80.4	85.4	62.9
c (Å)	91.5	89.6	98.9	36.5
β (°)	106.8	105.8	90	114.6
No. of reflections	650381	334177	875722	20088
No. of unique reflections	45147	110648	15513	5022
Redundancy	14.4(12.0)	3.0(2.5)	5.5(5.3)	4.0(3.0)
Completeness	98.9(92.5)	96.3(81.6)	99.6(96.2)	99.2(95.4)
R_{sym}[2](%)	7.8(30.9)	4.6(21.9)	8.0(24.7)	14.5(39.2)
I/σ(I)	40.1(6.6)	26.2(4.0)	18.6(4.1)	14.5(3.5)

[1] Merged with low resolution data from SeMet crystal.

[2] $R_{sym} = \Sigma\Sigma_i | I_i - \langle I \rangle | / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices hkl.

Numbers in parentheses represent the highest resolution shell.

Integration and scaling of all data was done using the HKL2000 suite of programs (*21*).

Structure Determination and Refinement

Selenium sites in the selenomethionine derivative CysM-CysO crystal were found using SOLVE (*22*), with 11 of the 14 possible selenium peaks located. Density modification (*23*) and automated model building (*24*) was done using RESOLVE. The remainder of the model was built using the graphics programs O (*25*) and Coot (*26*). Refinement was carried out using the CCP4 (*27*) and CNS (*28*) suite of programs. The selenomethionine derivative structure was then refined against the higher resolution native CysM-CysO data. The native CysM-CysO dataset had low completeness in the resolution range of 6 Å and below due to intensity overloads. To account for this, the native CysM-CysO dataset was merged with data corresponding to a resolution of 6 Å and below from the selenomethionine derivative crystal in Scalepack (*21*) for refinement and map calculations. In the final stages of refinement, the high resolution data alone were used to refine the model. All residues in all three molecules were modeled into the electron density with the exception of the N-terminal methionine residue of CysO and residues 126-129 in the CysM chain that does not interact with CysO ($CysM_B$). Six residues were modeled in alternate conformations: Leu169 from the CysM molecule that interacts with CysO ($CysM_A$), Cys99, Met190, Thr257 and Leu270 from $CysM_B$ and Thr9 from CysO.

The structure of $CysM_A$ was used as an initial model to refine against the K204A CysM data. Residues 212-226 and 315-323 from chain A and 313-326 from chain B were not modeled due to disorder in the electron density maps. The refined structure of CysO from the CysM-CysO complex was used as an initial model to refine the CysO only data. All but residues 91-93 were built into the model. The final

Table 3.2. Refinement Statistics

Sample	SeMet CysOM	Native CysOM	K204A CysM	Native CysO
No. of non-H atoms	5529	6512	4648	1290
No. of protein atoms	5150	5454	4443	1266
No. of water atoms	349	1028	160	24
No. of ligand atoms	30	30	45	0
$R_{working}$ (%)[3]	19.2	17.6	19.0	19.5
R_{free} (%)[4]	24.3	21.7	25.6	25.2
Ramachandran plot				
Most favored region (%)	89.7	91.7	89.4	89.0
Additionally allowed regions (%)	7.5	7.9	8.6	11.0
Generously allowed region (%)	1.9	0.3	1.6	0.0
Disallowed region (%)	0.9	0.0	0.4	0.0
R.M.S. deviations from ideal				
Bonds (Å)	0.021	0.026	0.021	0.019
Angles (°)	2.0	2.2	2.0	2.0

[3] $R_{working} = \Sigma||F_{obs}| - K|F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively

[4] R_{free} was calculated using 5% of all reflections that were excluded at all stages of refinement

refinement statistics for all datasets and the geometries and distances of all final models as evaluated by PROCHECK (*29*) are shown in Table 3.2.

Molecular Modeling of Intermediates

The program MacroModel, version 9.1 (*30, 31*) was used to model both the O-acetylserine (OAS)-PLP adduct **3**, the α-aminoacrylate intermediate **5** and intermediate **6**. All water molecules were removed from the structure and the C-terminus of CysO was converted to a thiocarboxylate and manually adjusted to point toward the active site. Hydrogen atoms were added where appropriate. The initial position of each intermediate was approximated by overlaying the structure of O-acetylserine sulfhydrylase from *Salmonella typhimurium* containing an external aldimine linkage to methionine (*32*) with CysM and superimposing their respective aldimine Cα and carboxylate atoms. Modeling was carried out using a 20 Å shell surrounding the OAS moiety of the adduct, with residues within a 6 Å shell allowed to move and the remaining residues frozen. Some of the key residues in the conformational search included 91-AGG-93 from CysO, and 184-GTTGT-188, Lys51, Ser265 and 322-WA-323 from CysM. Torsional rotation was allowed between all atoms within the PLP-OAS adduct connected by a single bond in the conformational search parameters. Modeling was done using the AMBER* force field (*33, 34*) and a distance dependent electrostatic treatment with a dielectic constant of 4.0. The TNCG minimization method was used for the energy minimization of structures generated by conformational search (*35*). Conformational searching with the α-aminoacrylate intermediate was done in the same way. The final models were manually adjusted to optimize the expected reaction geometry.

Figure Preparation

All figures of protein molecules and residues were generated in PyMOL (*36*).

Section 3.3 Results

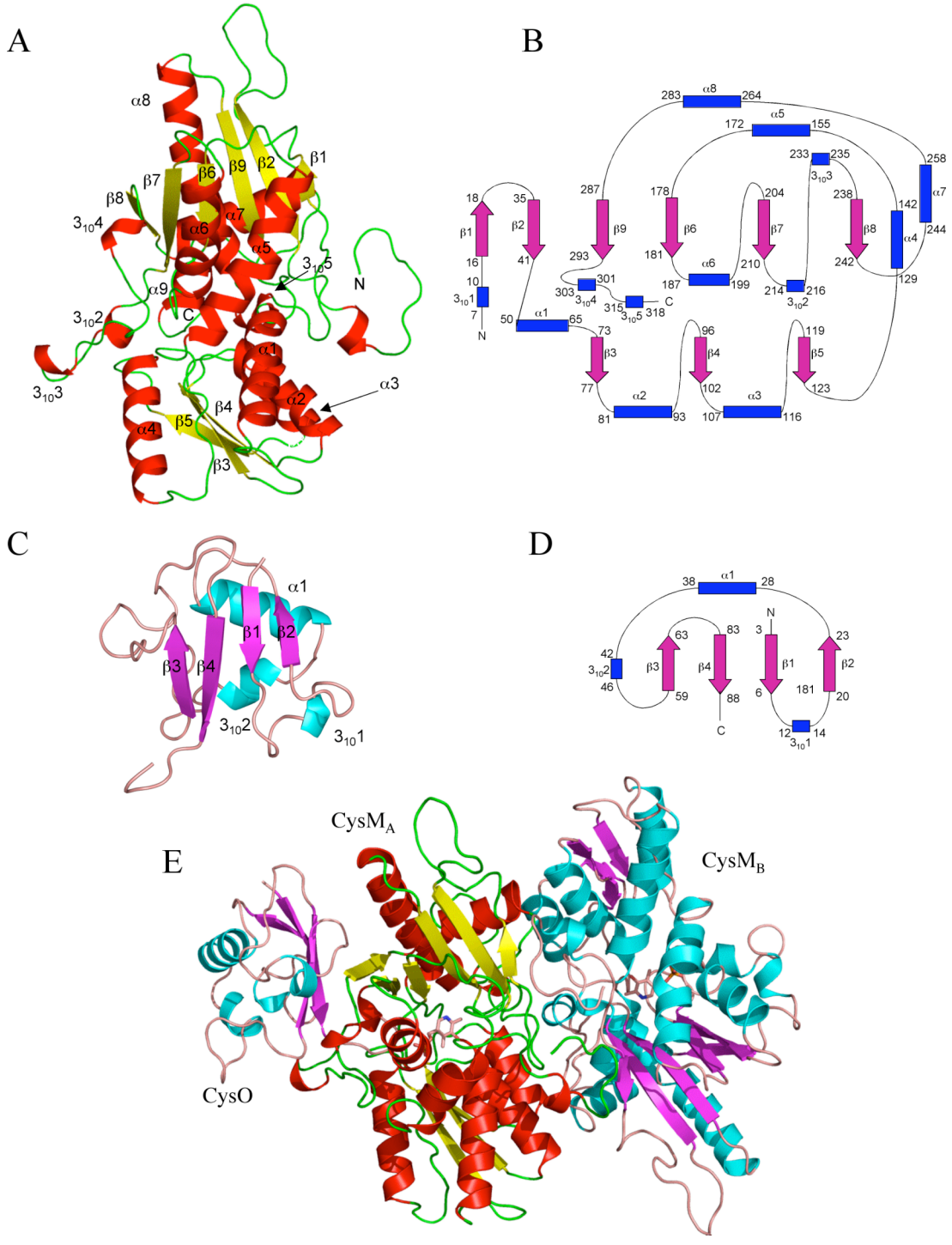
CysM Structure

The CysM molecule is composed of eight α-helices, five 3_{10}-helices and nine β-strands (Figure 3.2A and B). The protomer has a large and a small domain, each with an αβα sandwich fold. The large domain has a mixed β-sheet with β1↑β2↓β9↓β6↓β7↓β8↓ topology, while the smaller domain consists of a three-stranded parallel β-sheet comprised of strands β3, β4 and β5. This topology is consistent with other β-elimination enzymes (*37*). The interface between CysM molecules consists of loop regions with 60% of the residues being hydrophobic at the dimer interface. Other interactions include a salt bridge between Arg3 and Glu175 and a series of hydrogen bonding pairs that are listed in Table 3.3.

Table 3.3. Dimeric Interactions

A Chain	B Chain	H-bond Atoms
Thr2	Asp172	OH–Oδ
Tyr4	Leu16	NH–O
Tyr4	Gly18	O–NH
Asp5	Gln20	O–NHε
Leu92	Arg21	O–NH1
Gln111	Leu301	NHε–O
Leu115	His256	O–NHε
Tyr116	Gly259	OH–O
Asp172	Arg3	O–NH1
Glu175	Arg3	Oε–NH1
His256	Arg91	O–Nε

Figure 3.2. Structures of CysM and CysO. (A) CysM ribbon diagram. (B) CysM topology. (C) CysO ribbon diagram. (D) CysO topology diagram. (E) Ribbon diagram of the CysM-CysO complex. CysM bound to CysO ($CysM_A$) is color coded with red helices, yellow strands and green loops. CysO and CysM not bound to CysO ($CysM_B$) are color coded with cyan helices, magenta strands and salmon loops.


A
α8
β1
β2
β9
β6
β7
β8
α7
α6
α5
α9
C
N
α1
α2
α3
β4
β5
α4
β3
B
C
α1
β1
β2
β3
β4
D
E
CysM_A
CysM_B
CysO

CysO Structure

The CysO monomer is 93 residues in length and 9.6 kDa in mass. The secondary structure consists of two 3_{10} helices, a mixed four-stranded β-sheet with a β3↑β4↓β1↓β2↑ topology and one α-helix inserted between β2 and β3 (Figure 3.2C and D). This topology describes the β-grasp fold that is seen in the sulfur carrier proteins ThiS (*15*) and MoaD (*17*) as well as ubiquitin (*38*).

CysM-CysO Complex

The asymmetric CysM-CysO complex consists of a CysM dimer with a CysO molecule bound to one end ($CysM_A$ denotes the protomer interacting with CysO and $CysM_B$ denotes the unbound protomer) (Figure 3.2E). The $CysM_A$ and $CysM_B$ protomers show two main differences. First, the loop region spanning residues 211-237 in $CysM_A$ contains three 3_{10} helices, and extends away from the core of the structure to accommodate CysO binding. $CysM_B$ adopts a more compact structure with an α-helix spanning residues 217-222 and a single 3_{10} helix spanning residues 228-231. Second, the smaller domain comprising strands β3, β4 and β5 and helix α4 is shifted away from the core of the complex structure. Only the conformation of the CysO binding loop in $CysM_A$ allows for hydrogen bonds to form between the region connecting β5 and α4 in the smaller domain. The hydrogen bonds that contribute to this shift in $CysM_A$ are between the carboxylic acid side chain from Glu126 and the carbonyl oxygen atom of Ala218 and between the amide nitrogen atom of Gly217 and the carbonyl oxygen atom of Val216.

The comparison between CysO molecules from the complex structure and the CysO alone structure shows that the overall structure remains the same, with a root-mean-square deviation (r.m.s.d.) between 89 Cα carbon atoms of 0.5 Å. One main difference is that the C-terminus of CysO in the complex is well ordered through several hydrogen bonds with CysM. The absence of CysM binding leads to disorder

in the final three residues of CysO alone, which were not modeled into the structure. 25% of the accessible surface of CysO is buried in the complex as calculated by the protein-protein interaction server (*39*). Key hydrogen bonds between CysO and CysM are listed in Table 3.4. The side chain oxygen atom of $Asp65_{CysO}$ is hydrogen bonded to $His271_{A}$ through a bridging water molecule (subscripts CysO and A refer to CysO and $CysM_{A}$, respectively). A bridging water molecule is found donating hydrogen bonds to the carbonyl oxygen atoms of both $Gly92_{CysO}$ and $Leu183_{A}$, and two bridging water molecules form a hydrogen bonding network connecting the carbonyl oxygen atoms of $Val90_{CysO}$ and $Phe226_{A}$. Two salt bridges also occur at the protein-protein interface between $Arg12_{CysO}$ and $Glu214_{A}$, and $Asp65_{CysO}$ and $Arg211_{A}$.

Table 3.4. CysM-CysO Hydrogen Bonds

CysM	CysO	H-bond Atoms
Asn130	Ala89	NHδ–O
Asn130	Ala91	Oε–HN
Glu209	Tyr62	Oε–HO
Tyr217	Gly93	OH–O
Phe226	Ala89	NH–O
Arg239	Asp67	O–HN
Ser241	Asp65	OH–Oδ

Active Site

A PLP cofactor is found in each protomer of CysM covalently attached to Lys51 through a Schiff base linkage. Figure 3.3 illustrates key interactions between CysM and PLP. A glycine and threonine rich sequence binds the phosphate group, and is characteristic of PLP binding proteins (*40*). A large area of open space occupied only by water molecules spans the distance between Gly93 of CysO and the

PLP molecule. The distance from the Schiff base forming carbon atom of PLP to the closest C-terminal oxygen atom of CysO is 11.3 Å.

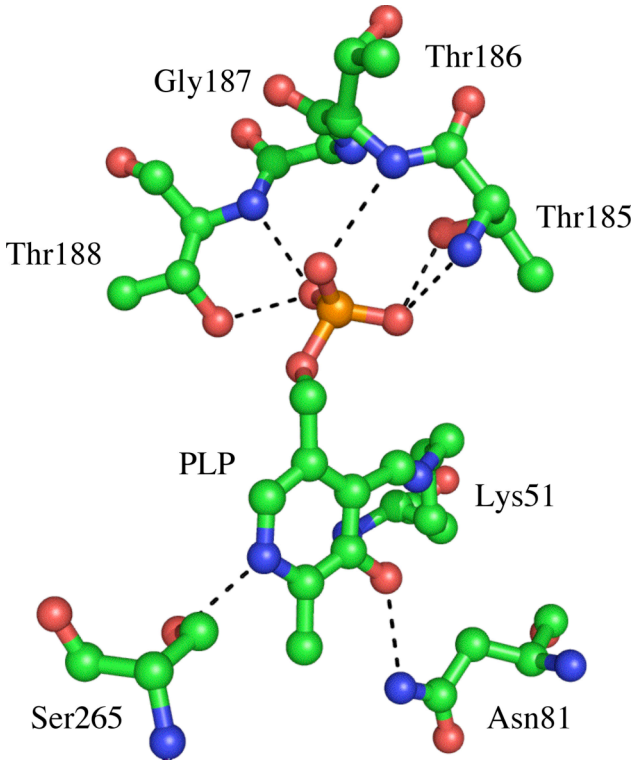


Figure 3.3. CysM active site with PLP coordinating residues labeled with green carbon atoms.

Molecular Modeling

Molecular modeling simulations using conformational searching followed by molecular mechanics based energy minimization, followed by manual oprimization, were carried out using MacroModel for both the PLP-OAS adduct **3** and the α-aminoacrylate intermediate **5** in the active site. In the PLP-OAS model (Figure 3.4A), the carboxylate moiety of OAS accepts a hydrogen bond from the amino side chain of Asn81. One oxygen atom of the acetate moiety accepts a hydrogen bond from the hydroxyl group of Tyr212 while the other accepts a hydrogen bond from the amide nitrogen atom of Ala323. In the α-aminoacrylate model (Figure 3.4B), the carboxylate moiety accepts hydrogen bonds from the hydroxyl group of Tyr212 and the amide nitrogen atom of Ala323 as well. The conformation of the C-terminus of CysO shows the sulfur atom at a distance of 3.7 Å from the β-carbon α-aminoacrylate intermediate **5**. Using the aminoacrylate as a starting point, a model for the subsequent intermediate **6** in which the carbon-sulfur bond has formed was generated (Figure 3.4C).

Section 3.4 Discussion

CysM and Structural Homologues

Structural homologues were identified using the DALI server (www.ebi.ac.uk/dali) and the top hits were selected for comparison (Table 3.5). The closest structural homolouges are β-elimination enzymes, which use OAS or phosphoserine as a substrate to make cysteine. CysK, the sulfide-dependent cysteine synthase, has r.m.s.d. values of 2.4 Å for 281 Cα carbon atoms from $CysM_A$ and 1.8 Å for 278 Cα carbon atoms from $CysM_B$. *S. typhimurium* O-acetylserine sulfhydrylase (OASS) has r.m.s.d. values of 2.5 Å for 287 Cα carbon atoms from $CysM_A$ and 1.9 Å for 283 Cα carbon atoms from $CysM_B$. The conformation of the CysO binding loop of CysM is the most significant difference between the structural homologues, which do not bind

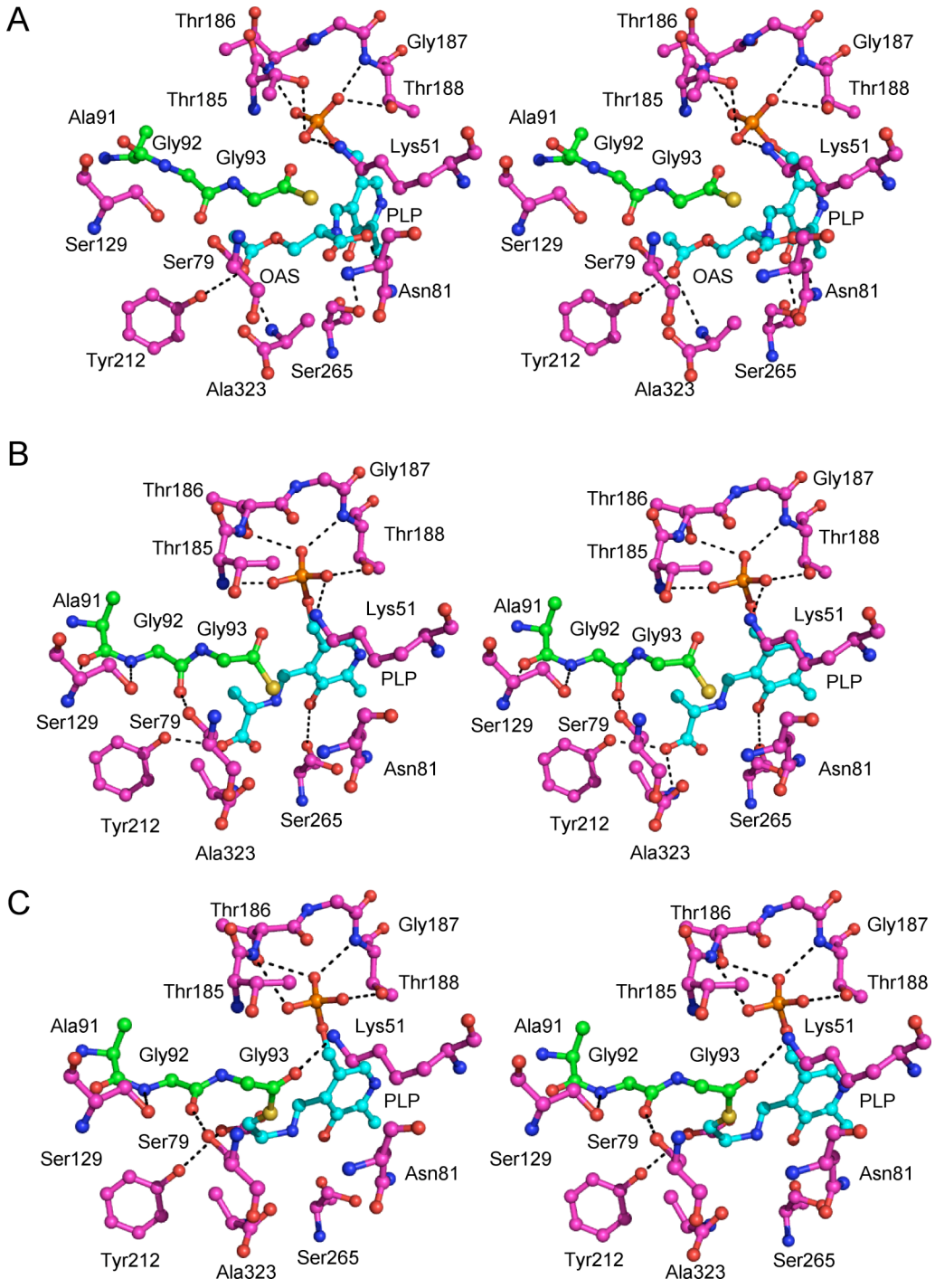
a sulfur-carrier protein. The conformation of the binding loop in $CysM_B$ is more similar to

Table 3.5. CysM Structural Homologues

PDB Code	Z-Score	Aligned Cα atoms	r.m.s.d.	Annotation
2Q3B	34.8	281	2.4	cysteine synthase
1OAS	34.2	287	2.5	O-acetylserine sulfhydrylase
1WKV	32.7	278	2.0	O-phosphoserine sulfhydrylase
2TYS	27.9	274	2.7	tryptophan synthase
1TDJ	25.9	271	2.9	threonine deaminase
1V71	24.7	267	3.2	serine racemase
1PWE	23.9	264	3.4	serine dehydratase
1F2D	22.9	262	2.7	1-aminocyclopropane-1-carboxylate deaminase
1E5X	21.7	269	3.6	threonine synthase
1KL7	20.7	269	3.5	threonine synthase

the structural homologues than in $CysM_A$, which results in the lower r.m.s.d. values. Another structural difference is that the C-terminal tail of CysM encompassing residues 300-323 contains two 3_{10} helices and points directly into the active site. Both the CysK and OASS structures lack this N-terminal extension.

Figure 3.4. Stereoviews of the active site models. (A) The PLP-OAS adduct **3**. (B) α-aminoacrylate intermediate **5**. (C) Intermediate **6** in which the C-S bond has formed. CysM residues are labeled with magenta carbon atoms, CysO residues are labeled with green carbon atoms and PLP and OAS or aminoacrylate moieties with cyan carbon atoms.


A
Thr186
Gly187
Thr188
Thr185
Ala91
Gly92
Gly93
Lys51
PLP
Ser129
Ser79
OAS
Asn81
Tyr212
Ala323
Ser265
Thr186
Gly187
Thr188
Thr185
Ala91
Gly92
Gly93
Lys51
PLP
Ser129
Ser79
OAS
Asn81
Tyr212
Ala323
Ser265
B
Gly187
Thr186
Thr188
Thr185
Ala91
Gly92
Gly93
Lys51
PLP
Ser129
Ser79
Asn81
Tyr212
Ala323
Ser265
Gly187
Thr186
Thr188
Thr185
Ala91
Gly92
Gly93
Lys51
PLP
Ser129
Ser79
Asn81
Tyr212
Ala323
Ser265
C
Thr186
Gly187
Thr188
Thr185
Ala91
Lys51
Gly92
Gly93
PLP
Ser79
Ser129
Asn81
Tyr212
Ala323
Ser265
Thr186
Gly187
Thr188
Thr185
Ala91
Lys51
Gly92
Gly93
PLP
Ser79
Ser129
Asn81
Tyr212
Ala323
Ser265

CysM and CysK Active Site

The PLP binding geometry of CysM and CysK are very similar (Figure 3.5A and B). The final two residues of CysM, Trp322 and Ala323, point toward the active site of the protein. A π-stacking interaction between Trp322 and Tyr212 helps to position the C-terminus. Residues 218-222, rather than the C-terminus, occupy this region in CysK. These residues correspond to the CysO binding loop of CysM based on sequence alignment. Additionally, there is a molecule of 2-methyl-2,4-pentanediol in CysK corresponding to the region of CysM spanning the C-terminus of CysO and the PLP cofactor. The C-terminus of the CysM K204A mutant also does not point to the active site. However, the CysO binding loop of one of the K204A protomers is seen ordered and has the same conformation as the corresponding residues in CysK.

CysO and Structural Homologues

Structural homologues were identified using the DALI server and selecting the top hits for comparison (Table 3.6). The β-grasp proteins complexed with a binding partner were also used for comparison. This included ThiS, MoaD and ubiquitin, which were compared to CysO through structural superposition. Each possesses the same topology, with most structural differences found in loop regions rather than secondary structural elements (Figure 3.6). CysO has insertions spanning residues 7-19 and 48-55, which are absent in the homologues. The exception is MoaD, which aligns well with residues 7-19 in CysO. Both CysO and MoaD form a key binding interaction with their respective binding partner in this region. The r.m.s.d. values between Cα positions in CysO are 2.6 Å for ThiS for 58 Cα carbon atoms, 2.8 Å for MoaD for 77 Cα carbon atoms and 2.6 Å for ubiquitin for 62 Cα carbon atoms.

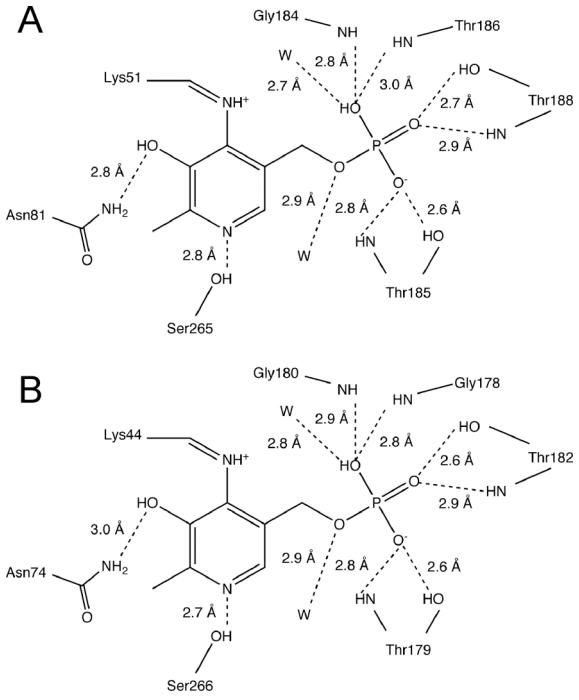


Figure 3.5. Schematic representation of CysM (A) and CysK (B) binding interactions with the PLP cofactor.

Table 3.6. CysO Structural Homologues

PDB Code	Z-Score	Aligned Cα atoms	r.m.s.d.	Annotation
1V8C	14.9	86	1.8	*Thermus thermophilus* – MoaD
1FMA	9.5	78	2.3	MoaD from MoaE–MoaD Complex
1JW9	8.9	77	2.8	MoaD from MoeB–MoaD Complex
2GLE	8.9	82	2.9	Neurabin SAM domain
1XO3	8.1	80	2.9	*Mus musculus* – Urm1
2BB5	7.9	78	2.6	Transcobalamin II
1C1Y	5.4	63	2.5	Ras-binding domain of c-Raf1
1ZUD	5.1	58	2.6	ThiS from ThiF–ThiS Complex
1TYG	4.7	59	3.2	ThiS from ThiG–ThiS Complex
1ZGU	4.4	62	2.6	Ubiquitin from MMS2–Ub Complex

CysM-CysO: Interactions and Conserved Residues

The features required for interactions with CysO come from three main sections of the CysM molecule. First is the CysO binding loop encompassing residues 211-237, which contributes the majority of CysO binding interactions. Second is helix α4, which is part of the smaller domain of the CysM molecule. Asn130 from this helix forms one of the key hydrogen bonds that stabilizes the C-terminal region of

CysO. Third, strand β8 forms hydrogen bonds with the loop that proceeds from the β3 strand of CysO.

In order to determine key CysM residues, sequence alignments were performed using both homologues that utilize sulfur carrier proteins and ones that incorporate sulfur directly from sulfide. Residues Asn81 and Ser265 that bind the PLP cofactor are exclusively conserved in both types. The glycine and threonine rich sequence that binds the phosphate group of the cofactor varies slightly in the position of glycine and

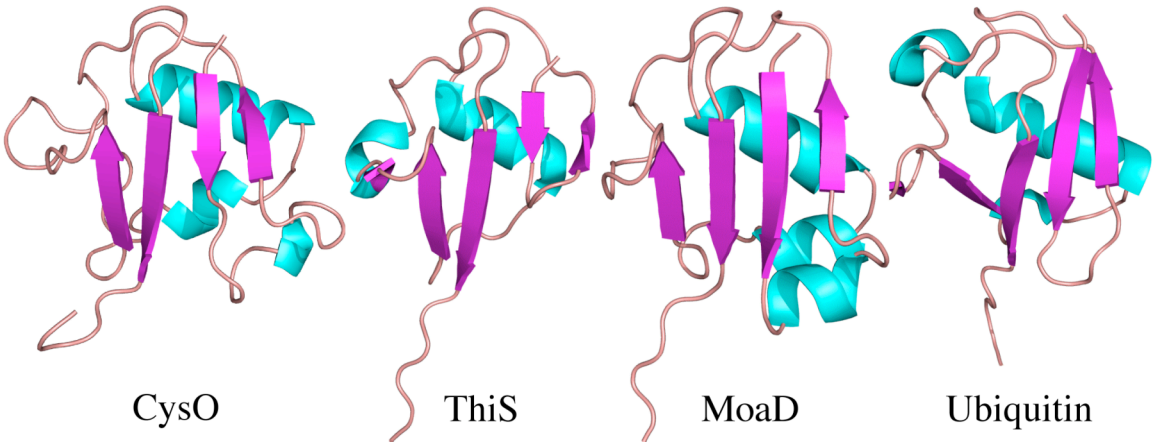


Figure 3.6. Ribbon diagrams of all known sulfur carrier proteins and ubiquitin.

threonine or serine residues, but overall the motif is conserved. A sequence alignment of 200 non-redundant CysM sequences shows that none of the residues on the CysM surface involved with CysO binding are conserved. When the alignment is restricted to the organisms *M. vanbaalenii*, *M. smegmatis*, *M. ulcerans* and *M. flavescens*, all of which have both CysO and CysM genes, the CysO binding residues are exclusively conserved. Conservation of the CysM-CysO binding interface is maintained within the mycobacterium genus, but shows greater variation, especially in CysM, for other species.

Comparison of the CysM-CysO Interface with Other β-grasp Protein Complexes

In order to determine if regions of CysM and CysO involved in protein-protein interactions are similar to other β-grasp protein complexes, the structures ThiF-ThiS (*15*) ThiG-ThiS (*16*), MoeB-MoaD (*17*), MoaE-MoaD (*10*) and ubiquitin in complex with the E2-like protein MMS2 (*41*) were compared in Figure 3.7. Table 3.7 lists several properties of the sulfur carrier proteins and ubiquitin at the protein-protein interface. The percent of hydrophobic residues at the interface ranges from 61-72% for the sulfur carrier protein examples, while the ubiquitin-MMS2 structure falls below this range at 58%. The number of hydrogen bonds between sulfur carrier proteins and their binding partners range from nine to 11, but ubiquitin-MMS2 structure has only five. This is due to ubiquitin having much less of its accessible surface area in close contact with its binding partner. In the sulfur carrier protein examples, 20-25% of the accessible surface area is in close contact with each binding partner. In ubiquitin, only 13% of the surface makes up the binding interface with MMS2.

Structural alignments were made by superimposing each sulfur carrier protein or ubiquitin complex using only CysO as the template. No structural similarities

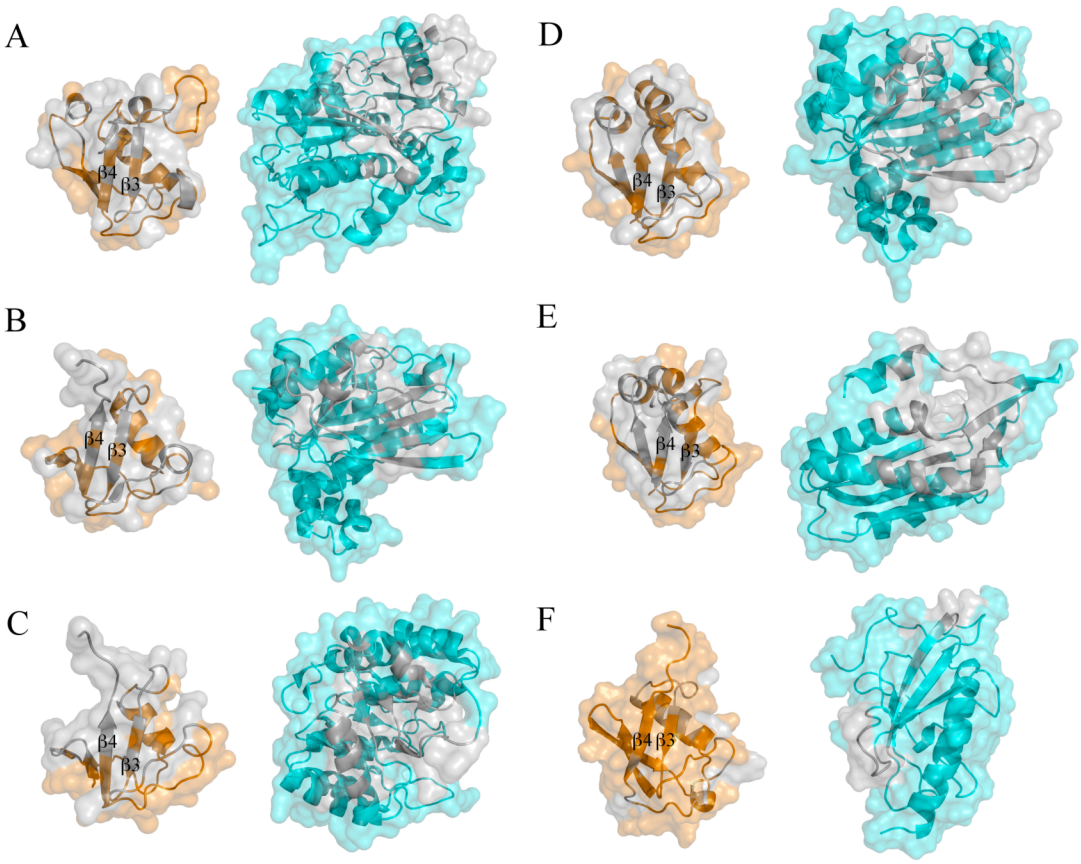


Figure 3.7. Ribbon and surface interface diagrams of CysM-CysO. (A) ThiF-ThiS complex. (B) ThiG-ThiS complex. (C) MoeB-MoaD complex. (D) MoaE-MoaD complex. (E) Ubiquitin-MMS2 complex. Sulfur carrier proteins and their respective binding partners are colored orange and light blue, respectively, in the areas that do not contribute to the binding interface. The silver colored regions in both proteins indicate the area that makes up the binding interface.

Table 3.7. Protein-Protein Interface Comparisons

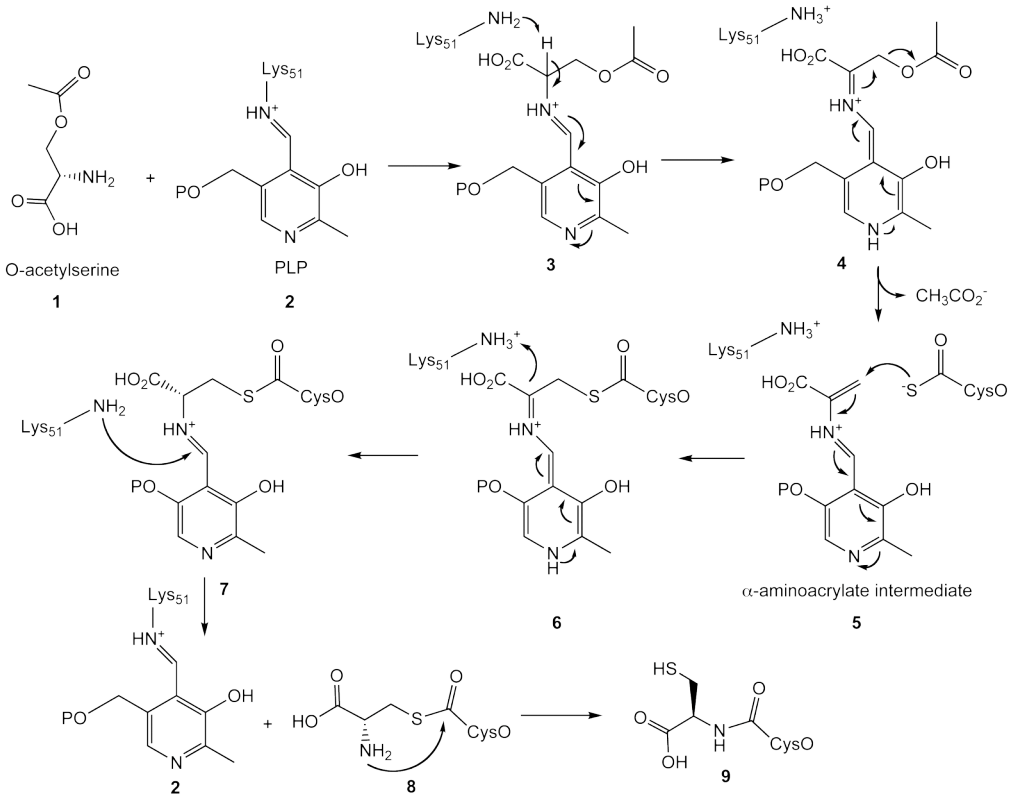
Protein Interface Parameter	CysO-CysM	ThiS-ThiF	ThiS-ThiG	MoaD-MoaB	MoaD-MoaE	Ub-MMS2
Interface Accessible Surface Area ($Å^2$)	1213.6	1045.9	1194.2	1113.2	985.1	632.4
% Interface Accessible Surface Area	24.6	22.7	24.2	23.5	19.7	13.4
% Polar Atoms in Interface	35.1	38.2	28.2	35.4	39.3	42.2
% Non-Polar Atoms in Interface	64.8	61.7	71.7	64.6	60.6	57.8
Hydrogen Bonds	11	14	11	9	13	5
Salt Bridges	1	1	0	1	0	0

between CysM, ThiG, MoaE or MMS2 are observed. However, two β-strands between CysM, ThiF and MoeB align when each sulfur carrier protein is overlaid. The strands β7 and β8 in CysM align with strands β6 and β7 in ThiF and MoeB despite the fact that CysM has a different fold. The regions in the sulfur carrier proteins and ubiquitin that make up most of the binding interface come from strands β3, β4 and the residues following β4 to the C-terminus. The ubiquitin complex

structure has the smallest interface with only residue 42 from strand β3 and residue 68 from β4 contributing to hydrogen bonds at the interface.

Mechanistic Implications

Based on the extensive mechanistic characterization of the sulfide-dependent cysteine synthase (*42*) a mechanistic proposal for the CysO/CysM-catalyzed reaction is outlined in Scheme 1. In this proposal, O-acetyl serine forms an imine with PLP, which then undergoes a deprotonation to give **4**. Acetate elimination followed by



Scheme 3.1

thiocarboxylate addition to the resulting aminoacrylate gives **6**. The reaction scheme is completed by a tautomerization to **7**, and a transimination reaction with Lys51 to release thioester **8**, which undergoes an N/S acyl shift to give cysteinyl-CysO.

The structural studies reported here are consistent with this mechanistic proposal. PLP is bound at the active site via an imine with Lys51. The model of the

enzyme substrate complex shown in Figure 3.4A shows that the carboxy terminal sulfur of CysO thiocarboxylate is suitably positioned for addition to the β-carbon of O-acetyl serine (separation ~3.7 Å), that the acetyl group is activated as a leaving group by hydrogen bonding to an oxyanion hole comprising of the hydroxyl of Tyr212, and that the amide NH of Ala323. Lys51 is the likely base mediating the deprotonation that converts **3** to **4**. The model of the aminoacrylate intermediate shown in Figure 3.4B is also consistent with our mechanistic proposal. In this model, the carboxylate of the aminoacrylate **5** has moved into the oxyanion hole vacated by the departure of acetate further activating the aminoacrylate for addition of the thiocarboxylate sulfur.

Section 3.5 Conclusion

The structure of CysM-CysO is an example of a β-elimination enzyme in complex with a sulfur carrier protein used to generate the amino acid cysteine. CysM is able to bind CysO using a binding loop that is ordered differently in structural comparators and when CysO is unbound. The binding loop contains residues that are only conserved among other species within the mycobacterium genus that also utilize a sulfur carrier protein for catalysis. The overall fold of CysO is similar to that of other sulfur carrier proteins and ubiquitin. Two β-strands at the interface of CysO and CysM are structurally aligned when compared to the E1 adenyltransferase proteins ThiF and MoaD. Modeling studies show that the C-terminus of CysO is able to reach the β carbon of the α-aminoacrylate intermediate in a conformation favorable for nucleophilic addition.

BIBLIOGRAPHY

1. Burns, K. E., Baumgart, S., Dorrestein, P. C., Zhai, H., McLafferty, F. W., and Begley, T. P. (2005) Reconstitution of a new cysteine biosynthetic pathway in *Mycobacterium tuberculosis*, *J. Am. Chem. Soc. 127*, 11602-11603.
2. Kusner, D. J. (2005) Mechanisms of mycobacterial persistence in tuberculosis, *Clin. Immunol. 114*, 239-247.
3. Li, Z., Kelley, C., Collins, F., Rouse, D., and Morris, S. (1998) Expression of katG in *Mycobacterium tuberculosis* is associated with its growth and persistence in mice and guinea pigs, *J. Infect. Dis. 177*, 1030-1035.
4. Russell, D. G. (2003) Phagosomes, fatty acids and tuberculosis, *Nat. Cell. Biol. 5*, 776-778.
5. Vachula, M., Holzer, T. J., and Andersen, B. R. (1989) Suppression of monocyte oxidative response by phenolic glycolipid I of *Mycobacterium leprae*, *J. Immunol. 142*, 1696-1701.
6. Vergne, I., Chua, J., Singh, S. B., and Deretic, V. (2004) Cell biology of *Mycobacterium tuberculosis* phagosome, *Annu. Rev. Cell. Dev. Biol. 20*, 367-394.
7. Manganelli, R., Voskuil, M. I., Schoolnik, G. K., Dubnau, E., Gomez, M., and Smith, I. (2002) Role of the extracytoplasmic-function sigma factor sigma(H) in *Mycobacterium tuberculosis* global gene expression, *Mol. Microbiol. 45*, 365-374.
8. Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., and Cook, W. J. (1985) Three-dimensional structure of ubiquitin at 2.8 Å resolution, *Proc. Natl. Acad. Sci. U. S. A. 82*, 3582-3585.
9. Taylor, S. V., Kelleher, N. L., Kinsland, C., Chiu, H. J., Costello, C. A., Backstrom, A. D., McLafferty, F. W., and Begley, T. P. (1998) Thiamin

biosynthesis in *Escherichia coli*. Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation, *J. Biol. Chem. 273*, 16555-16560.

10. Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V., and Schindelin, H. (2001) Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation, *Nat. Struct. Biol. 8*, 42-46.
11. Godert, A. M., Jin, M., McLafferty, F. W., and Begley, T. P. (2007) Biosynthesis of the thioquinolobactin siderophore: an interesting variation on sulfur transfer, *J. Bacteriol. 189*, 2941-2944.
12. Xi, J., Ge, Y., Kinsland, C., McLafferty, F. W., and Begley, T. P. (2001) Biosynthesis of the thiazole moiety of thiamin in *Escherichia coli*: identification of an acyldisulfide-linked protein--protein conjugate that is functionally analogous to the ubiquitin/E1 complex, *Proc. Natl. Acad. Sci. U S A 98*, 8513-8518.
13. Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. (1999) Thiamin biosynthesis in prokaryotes, *Arch. Microbiol. 171*, 293-300.
14. Leimkuhler, S., Wuebbens, M. M., and Rajagopalan, K. V. (2001) Characterization of *Escherichia coli* MoeB and its involvement in the activation of molybdopterin synthase for the biosynthesis of the molybdenum cofactor, *J. Biol. Chem. 276*, 34695-34701.
15. Lehmann, C., Begley, T. P., and Ealick, S. E. (2006) Structure of the *Escherichia coli* ThiS-ThiF complex, a key component of the sulfur transfer system in thiamin biosynthesis, *Biochemistry 45*, 11-19.

16. Settembre, E. C., Dorrestein, P. C., Zhai, H., Chatterjee, A., McLafferty, F. W., Begley, T. P., and Ealick, S. E. (2004) Thiamin biosynthesis in *Bacillus subtilis*: structure of the thiazole synthase/sulfur carrier protein complex, *Biochemistry 43*, 11647-11657.

17. Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V., and Schindelin, H. (2001) Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex, *Nature 414*, 325-329.

18. Godert, A. M., Jin, M., McLafferty, F. W., and Begley, T. P. (2007) Biosynthesis of the thioquinolobactin siderophore: an interesting variation on sulfur transfer, *J. Bacteriol. 189*, 2941-2944.

19. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248-254.

20. Derewenda, Z. S. (2004) Rational protein crystallization by mutational surface engineering, *Structure 12*, 529-535.

21. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

22. Terwilliger, T. C., and Berendzen, J. (1999) Automated structure solution for MIR and MAD (http://www.solve.lanl.gov), *Acta Crystallogr. D 55*, 849-861.

23. Terwilliger, T. C. (2000) Maximum likelihood density modification, *Acta Crystallogr. D 56*, 965-972.

24. Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension, *Acta Crystallogr. D 59*, 38-44.

25. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A 47*, 110-119.

26. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D 60*, 2126-2132.

27. Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography, *Acta. Crystallogr. D 50*, 760-763.

28. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D 54*, 905-921.

29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

30. Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., and Still, W. C. (1990) MacroModel - an Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics, *J. Comput. Chem. 11*, 460-467.

31. MacroModel, v., Schrodinger, LLC, New York, NY, 2005.

32. Burkhard, P., Tai, C. H., Ristroph, C. M., Cook, P. F., and Jansonius, J. N. (1999) Ligand binding induces a large conformational change in O-acetylserine sulfhydrylase from Salmonella typhimurium, *J. Mol. Biol. 291*, 941-953.

33. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins, *J. Am. Chem. Soc. 106*, 765-784.

34. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986) An all atom force field for simulations of proteins and nucleic acids, *J. Comput. Chem. 7*, 230-252.

35. Ponder, J. W., and Richards, F. M. (1987) An Efficient Newton-like Method for Molecular Mechanics Energy Minimization of Large Molecules, *J. Comput. Chem. 8*, 1016-1024.

36. DeLano, W. L. (2002) The PyMOL User's Manual.

37. Burkhard, P., Rao, G. S., Hohenester, E., Schnackerz, K. D., Cook, P. F., and Jansonius, J. N. (1998) Three-dimensional structure of O-acetylserine sulfhydrylase from *Salmonella typhimurium*, *J. Mol. Biol. 283*, 121-133.

38. Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., and Cook, W. J. (1985) Three-dimensional structure of ubiquitin at 2.8 Å resolution, *Proc. Nat.l Acad. Sci. U S A 82*, 3582-3585.

39. Jones, S., and Thornton, J. M. (1995) Protein-protein interactions: a review of protein dimer structures, *Prog. Biophys. Mol. Biol. 63*, 31-65.

40. Momany, C., Ghosh, R., and Hackert, M. L. (1995) Structural motifs for pyridoxal 5´-phosphate binding in decarboxylases: an analysis based on the crystal structure of the Lactobacillus 30a ornithine decarboxylase, *Protein Sci. 4*, 849-854.

41. Lewis, M. J., Saltibus, L. F., Hau, D. D., Xiao, W., and Spyracopoulos, L. (2006) Structural basis for non-covalent interaction between ubiquitin and the ubiquitin conjugating enzyme variant human MMS2, *J. Biomol. NMR 34*, 89-100.

42. Tai, C. H., and Cook, P. F. (2001) Pyridoxal 5´-phosphate-dependent α,β-elimination reactions: mechanism of O-acetylserine sulfhydrylase, *Acc. Chem. Res. 34*, 49-59.

CHAPTER FOUR

CRYSTAL STRUCTURE OF A PURINE NUCLEOSIDE PHOSPHORYLASE (TM1596) FROM THERMOTOGA MARITIMA AT 2.01 Å RESOLUTION

Section 4.1 Introduction

The crystal structure of a trimeric purine nucleoside phosphorylase (PNP) from *Thermotoga maritima* (TmPNP; Tm1596) has been solved to 2.01 Å using X-ray crystallography. The structure reveals a bound guanine base surrounded by residues that are highly conserved among PNP orthologs. PNP enzymes have been well characterized biochemically (*1*), and X-ray crystal structures have been solved in both eukaryotic and prokaryotic systems (*2*). The structure of a PNP enzyme from a thermophilic organism provides insight into the thermostability of the enzyme when compared to a mesophilic counterpart, which in this communication will be the X-ray crystal structure of PNP from *Bos Taurus* (bPNP). Recent review articles discuss some of the factors which contribute to thermostability, and how these compare to mesophilic enzymes that catalyze the same reactions (*3, 4*). In this communication, we will highlight the structural characteristics that differ between these two proteins and how that contributes to the greater thermostability of TmPNP over bPNP.

Section 4.2 Materials and Methods

Protein crystals were grown and data collected by the Joint Center for Structural Genomics. Data collection statistics can be seen in Table 4.1.

The structure of TmPNP was solved using the method of molecular replacement with bPNP (PDB code 1PBN) as the molecular replacement search model. Cross-rotational and translation searching were performed using CNS (*5*). Rigid body refinement and simulated annealing were performed before calculating the $2F_o$-F_c and simulated annealing composite omit electron density maps. Model

Table 4.1. Data collection and Refinement Statistics

Resolution (Å)	2.01
Space group	P1
a, b, c	46.48 Å, 74.60 Å, 74.59 Å
α, β, γ	117.34°, 100.95°, 100.71°
No. of reflections	203633
No. of unique reflections	54061
Redundancy	3.8 (3.4)
Completeness (%)	97.7 (93.8)
R_{sym} [a] (%)	10.6 (52.2)
I/σ(I)	12.9 (2.6)
No. of non-H atoms	5969
No. of protein atoms	5600
No. of water atoms	335
No. of ligand atoms	34
R factor (%) [b]	20.4
R_{free} (%) [c]	24.0
Ramachandran plot	
Most favored region (%)	
Additionally allowed regions (%)	
Generously allowed region (%)	
Disallowed region (%)	
r.m.s.d. from ideal	
Bonds (Å)	0.017
Angles (°)	1.615

Table 4.1. con't Data collection and Refinement Statistics

[a] $R_{sym} = \Sigma\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma\langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices hkl.

[b] R factor $= \Sigma \mid |F_{obs}| - K|F_{calc}| \mid / \Sigma_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively

[c] R_{free} was calculated using 5% of all reflections that were excluded at all stages of refinement.

Numbers in parentheses are from the highest resolution shell.

building was done in O (*6*). Three-fold NCS averaging was used throughout initial refinement to improve electron density map calculations. The structure was refined using CNS (6). Final refinement statistics can be seen in Table 4.1.

All figures were made using PyMol (*7*) and ChemDraw®. Atomic coordinates and structure factors for TmPNP have been deposited into the Protein Data Bank (PDB) under the code 1VMK.

Section 4.3 Results and Discussion

The structure of TmPNP is very similar to that of bPNP, and has sequence similarity of 63% and identity of 44%. The TmPNP monomer consists of nine β-strands, seven α-helices and two 3_{10} helices (Figure 4.1). The β-strands can be divided up into two regions, one that makes a β-sheet and another that makes up a distorted β-barrel. The β-sheet is arranged with the topology β2↓β3↑β4↓β1↓ with helices α1 and α3 on one side of the sheet. The other side is solvent exposed up to the point were the β-sheet domain meets with the β-barrel. The α7 helix contains 27 residues and crosses over at this point between strands of the β-sheet and the β-barrel. The β-barrel contains two long strands, β5 and β9, which are 10 and 11 residues, respectively. Strand β5 forms parallel sheet like hydrogen-bonding contacts along β1 for the first five residues, and then twists into the β-barrel motif with the final five residues. The poor electron

density between residues 28-30 corresponds to the final turn of α2, which is not rendered in Figure 4.1. The numbering of the helices has been retained, however, to maintain the previously established topological convention (*2*).

The trimeric structure is characteristic of PNP enzymes seen in eukaryotes, although trimeric PNPs have also been seen in bacteria such as *Escherichia coli*, *Bacillus subtilis* and *Bacillus stearothermophilus (2)* (Figure 4.2A). One molecular interface between monomers consists of a long loop region spanning residues 132-159 that interacts with residues 177-188. An additional loop at this interface spans residues 77-88 and contains one 3_{10} helix. The other significant interface between

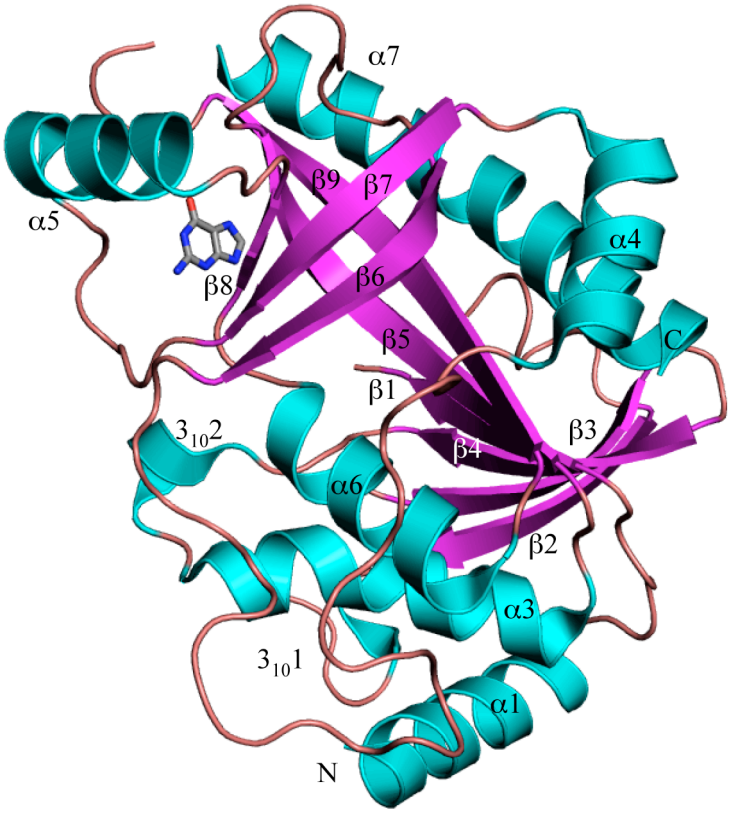


Figure 4.1. Secondary structural elements of the TmPNP monomer with guanine modeled into the active site. Strands are labeled in magenta, helices in blue and carbon atoms of guanine in grey.

monomers is seen along the three-fold axis surrounded by α5 from each monomer. The side of the trimeric structure containing the loop regions has the appearance of a bowl with the sides composed of the four-stranded mixed β-sheet and helix α1, and the base at the center of the three-fold axis. The opposite side of the trimer is flat with the center composed of the three α5 helices and the base of the β-barrel.

Each active site in TmPNP contains the purine guanine which is coordinated by exclusively conserved residues. The catalytic mechanism of PNP is facilitated by an asparagine residue, here Asn228, that stabilizes the negative charge that builds on the N7 nitrogen during catalysis (*1*). The N1 and N2 atoms of guanine are coordinated by the conserved residue Glu186. N9 forms a hydrogen bond with Ala111 and a water molecule, while π-stacking interactions with the ring are stabilized by the aromatic residue Tyr185 in a T-shaped configuration. Though Tyr185 is not exclusively

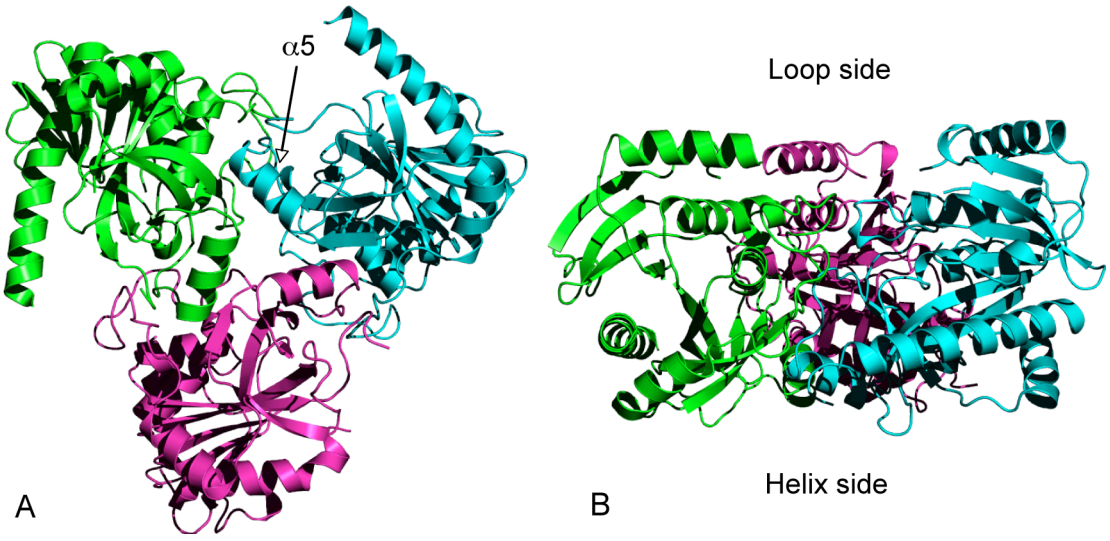
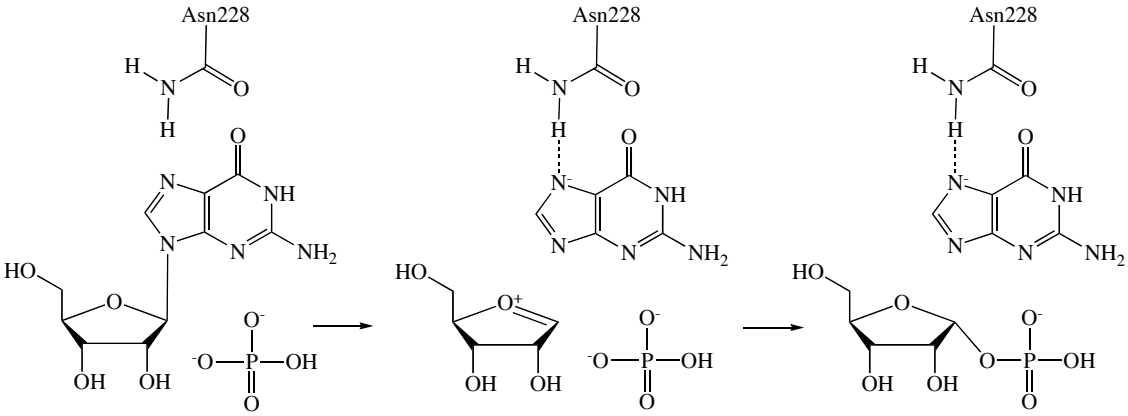


Figure 4.2. A) TmPNP trimer looking down the three-fold axis. B) Side view of the trimer rotated by 90° about the y-axis in figure A illustrating the loop and helix faces.

conserved, this position is conserved as either tyrosine or phenylalanine in order to conserve the π-π stacking interaction. These residues facilitate the catalytic mechanism that generates a free base and ribose 1-phosphate (Scheme 4.1).



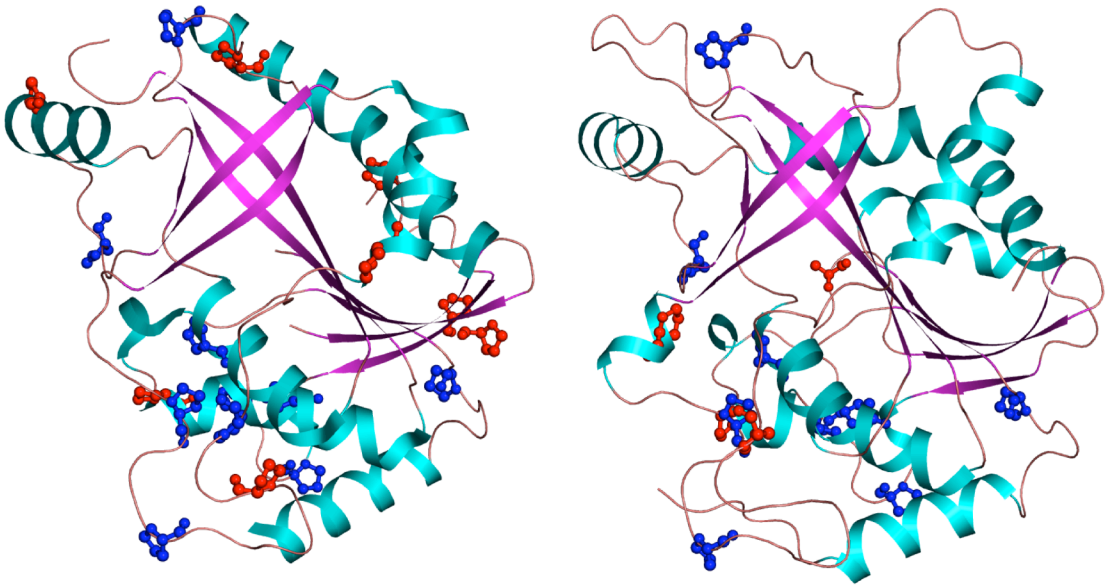
Scheme 4.1. PNP mechanism

Comparing thermophilic proteins to their mesophilic counterparts can provide insight into the fundamental nature of protein stability and folding in general. The structure of bPNP with both 9-deazahypoxanthine and ribose 1-phosphate bound (PDB code: 1A9T) serves as an excellent comparator for TmPNP since both proteins are the trimeric form of PNP, were solved to the same resolution, and have a structurally similar purine base bound in the active site.

Some of the key differences that have been implicated in contributing to thermostability are the number of proline residues, the presence of intramolecular disulfide bridges (*8*), number of salt bridges, hydrogen bonds, accessible surface area, increased polar surface area, hydrophibicity, fewer loop regions and greater compactness (*3*). Many of these factors are interrelated, and direct comparison of the X-ray crystal structures of thermophilic and mesophilic orthologs gives a specific, atomic resolution explanation for the differences.

TmPNP contains 19 proline residues and 11 salt bridges, while bPNP contains 13 proline residues and only six salt bridges. Figure 4.3 depicts the location of each proline residue in both bPNP and TmPNP. Nine of the 13 proline residues in bPNP are structurally aligned with those in TmPNP.

Another variance in residues is that thermophilic proteins tend to have fewer cysteine, asparagine and glutamine residues, as these residues tend to undergo oxidation and damage the protein at high temperatures (*9*). To confirm this, the number of cysteine, asparagine and glutamine residues was compared between the two. There is a significantly larger number of each residue in the bovine structure, with the number of cysteine, asparagine and glutamine residues at four, 14 and 15, respectively compared to TmPNP at two, seven and three, respectively. Intermolecular stability also plays a critical role in the formation of the biologically active unit. The interface of each monomeric unit must bind tightly to its neighbor so that at high temperatures the quaternary structure is not disrupted. Table 4.2 contains a list of properties generated by the protein-protein interaction server (*10*) and Spock (*11*), detailing binding interactions which stabilize the multimeric complex in both TmPNP and bPNP. The bovine intramolecular interface harbors more numerous binding features such as hydrogen bonds (nine in bPNP compared to seven in TmPNP) and significantly more bridging water molecules (11 in bPNP compared to four in TmPNP). The percent of non-polar atoms at this interface, with 67% in TmPNP and 60% in bPNP, excludes more water molecules in the TmPNP structure. This implies a more tightly packed trimeric structure that excludes water molecules from the interface due to the presence of hydrophobic residues. This is confirmed by comparing the van der Waals volume and surface area between bPNP



T. Maritima PNP *Bos taurus* PNP

Figure 4.3.: Proline residues in bPNP and TmPNP monomers shown with structurally conserved proline residues in blue and non-conserved proline residues in red.

Table 4.2. Intermolecular properties

Property	bPNP	TmPNP
Interface accessible surface area (Å^2)	1223	915
No. of Hydrogen Bonds	9	7
% Polar Atoms at Interface	40	33
% Non-Polar Atoms at Interface	60	67
Bridging Water Molecules	11	4
Van der Waals Surface Area (Å^2)	85813	77530
Van der Waals Volume (Å^3)	60874	53789
Ratio of S.A to Volume	1.41	1.44

Table 4.3. Monomeric properties

Property	bPNP	TmPNP	%difference[1]
No. of proline residues	13	19	37.5
No. of residues	289	277	4.2
No. salt bridges	6	11	58.8
Van der Waals Surface Area ($Å^2$)	28525	23308	20.1
Van der Waals Volume ($Å^3$)	20183	16828	18.1

1. $$\frac{Abs(X_1 - X_2)}{(X_1 + X_2)/2} \times 100$$

and TmPNP (Table 4.3). TmPNP has a significantly smaller surface area and volume compared to bPNP, as shown by the % difference in surface area and volume of 20.1% and 18.1%, respectively. The conclusion is that the factors stabilizing the trimeric TmPNP structure include having a more compact trimer as a result of the greater number of hydrophobic residues at the interface of adjacent monomers combined with the lower degree of intramolecular solvation.

The active site of TmPNP contains a guanine base, so the structurally similar 9-deazahypoxanthine bPNP structure was used for comparison. Figure 4.4 illustrates the active site of bPNP and TmPNP with their respective ligands bound and hydrogen bonding interactions shown. Residues conserved among PNP enzymes include a catalytic asparagine residue and a glutamic acid that stabilizes the N1 and N2 nitrogen atoms of guanine. In both cases there is a short hydrogen bond between the OE1 oxygen atom of glutamic acid and the N1 atom, with distances of 2.69 Å and 2.81 Å in TmPNP and bPNP, respectively. An aromatic residue is observed above the plane of the base, in a T-stacking π-stacking formation. This residue is Tyr185 in TmPNP and

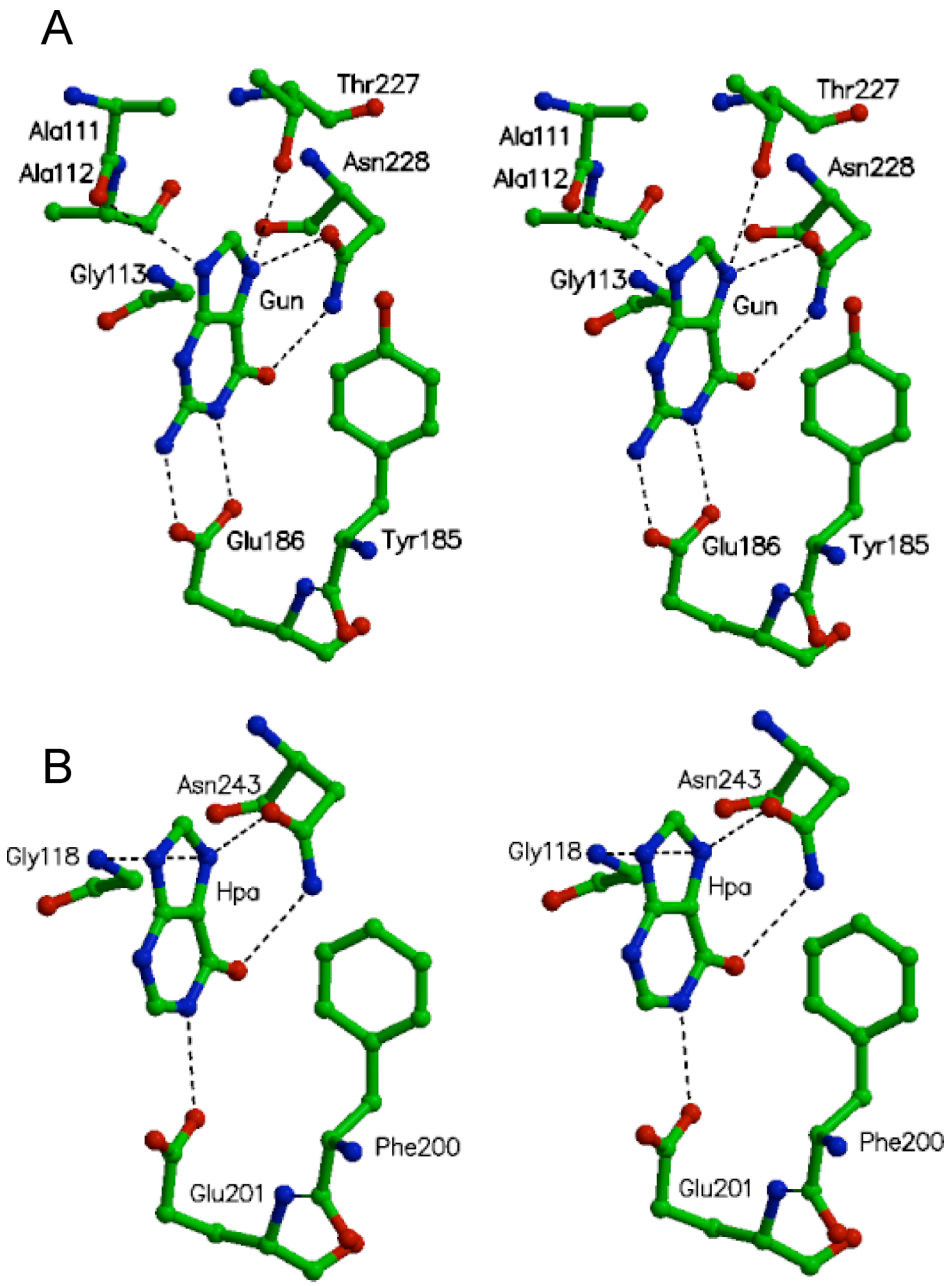


Figure 4.4. Stereoview of the active site residues of A) TmPNP with bound guanine (GUN) and B) bPNP with bound 9-deazahypoxanthine (HPA) and ribose 1-phosphate (R1P). Hydrogen bonds are shown as dashed black lines.

Phe200 in bPNP, but the aromaticity in this position is confirmed by multiple sequence alignment throughout the family of PNP enzymes.

The bovine structure also contains ribose 1-phosphate bound by several conserved residues. Though the TmPNP structure lacks this phosphorylated sugar, overlaying the structures gives information about the geometry of the conserved residues required to do so. Figure 4.4B shows the same structurally aligned residues in the TmPNP structure with the exception of His240 which is not located near the TmPNP active site. This corresponds to the most significant structural difference between the two proteins, in that α7, the final helix in both structures, diverges with the bend in the bPNP structure occurring at the conserved residue Ala267. This bend positions the imadazole ring of His257 in bPNP to make a hydrogen bond with the C5′-OH group of ribose 1-phosphate. Figure 4.5 shows this divergence as well as the region corresponding to residues 234-238 in TmPNP that were not ordered in the electron density (residues 234-235 are not ordered in chain C). This bend is also seen in bPNP with hypoxanthine bound but lacking ribose 1-phosphate (PDB code: 1A9Q). The TmPNP structure lacks ribose 1-phosphate, and the final helix is not bent but moved away from the active site.

Comparing the structures of orthologous proteins from mesophiles and thermophiles offers insights into what fundamentally stabilizes intramolecular and intermolecular interactions within a protein. The main feature, observed in the pair highlighted in this communication, which stabilizes the oligomeric state of the active enzyme is a higher percentage of non-polar residues than polar residues at the interface of adjacent monomers. This results in fewer hydrogen bonding contacts and a lower degree of solvation at the interface. The stability of the oligomeric state of proteins at high temperatures appears to have a large contribution from the unfavorable entropic effects of breaking the binding interactions seen at the interface

and exposing hydrophibic residues to solvent. An interface with a large percentage of polar residues would be more easily solvated, thus leading to more bridging water molecules. Since water molecules are dynamic in nature, as opposed to residues which are fixed in place by the topology of the protein, this can weaken the overall intramolecular binding interaction in the oligomeric protein as ambient temperatures begin to approach the boiling point of water.

Comparisons between bPNP and TmPNP monomers show what is classically believed to be a set of characteristics that differ between mesophilic and thermophilic proteins. TmPNP contains a larger number of proline residues, fewer asparagine, glutamine and cysteine residues, and more salt bridges.

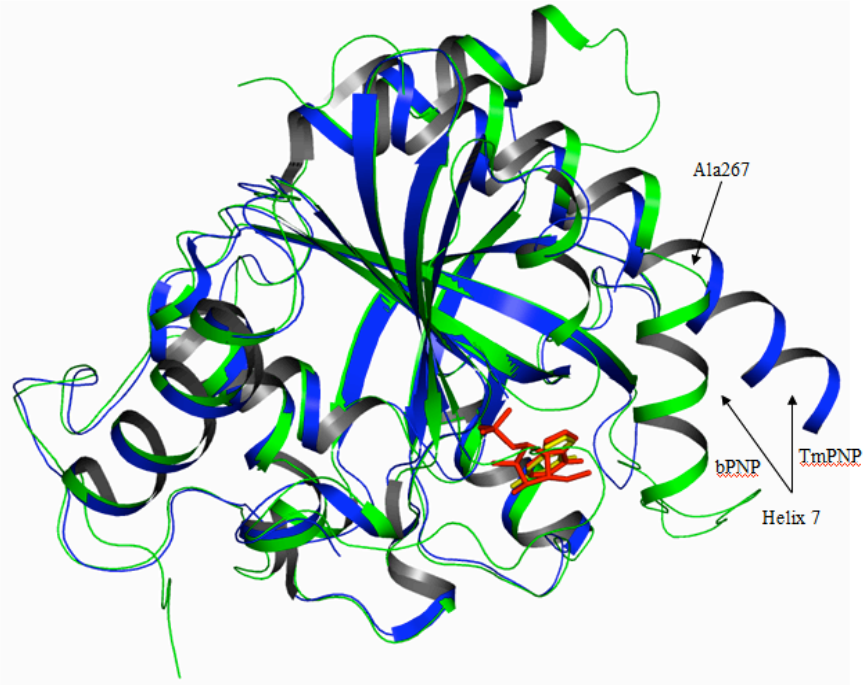


Figure 4.5. Overlaid structure of bPNP and TmPNP illustrating the divergences in the final helix, α7. bPNP is rendered in green, TmPNP in blue, guanine in yellow and both 9-deazahypoxanthine and ribose 1-phosphate in red.

BIBLIOGRAPHY

(1) Erion, M. D., Stoeckler, J. D., Guida, W. C., Walter, R. L., and Ealick, S. E. (1997) Purine nucleoside phosphorylase. 2. Catalytic mechanism. *Biochemistry 36*, 11735-48.

(2) Pugmire, M. J., and Ealick, S. E. (2002) Structural analyses reveal two distinct families of nucleoside phosphorylases. *Biochem. J. 361*, 1-25.

(3) Razvi, A., and Scholtz, J. M. (2006) Lessons in stability from thermophilic proteins. *Protein Sci. 15*, 1569-78.

(4) Sadeghi, M., Naderi-Manesh, H., Zarrabi, M., and Ranjbar, B. (2006) Effective factors in thermostability of thermophilic proteins. *Biophys. Chem. 119*, 256-70.

(5) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D 54*, 905-21.

(6) Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A 47 (Pt 2)*, 110-9.

(7) DeLano, W. L. (2002) The PyMOL User's Manual.

(8) Zhang, Y., Porcelli, M., Cacciapuoti, G., and Ealick, S. E. (2006) The crystal structure of 5'-deoxy-5'-methylthioadenosine phosphorylase II from *Sulfolobus solfataricus*, a thermophilic enzyme stabilized by intramolecular disulfide bonds. *J. Mol. Biol. 357*, 252-62.

(9) Tomazic, S. J., and Klibanov, A. M. (1988) Mechanisms of irreversible thermal inactivation of Bacillus α-amylases. *J. Biol. Chem. 263*, 3086-91.

(10) Jones, S., and Thornton, J. M. (1995) Protein-protein interactions: a review of protein dimer structures. *Prog. Biophys. Mol. Biol. 63*, 31-65.

(11) Christopher, J. A. (1998), SPOCK, Texas A&M University.

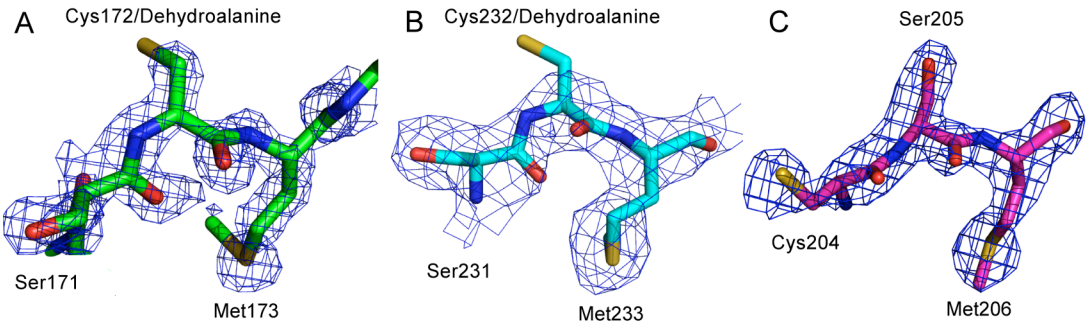
APPENDIX A

SUMMARY

The work presented here encompasses a wide range of interesting biological problems through the use of X-ray crystallographic methods. An atomic resolution structure of any enzyme is necessary to fully understand the nature of catalysis. The binding site geometry determines how the substrate not only interacts with the protein to fix it in place, but it also the chemistry through the precise positioning of atomic orbitals. Reactions that would never take place in solution are readily catalyzed by enzymes since their specificity is determined by their individual structure. As presented here, nature has devised a way to solve some complicated biological problems through highly creative means.

The structure of Thi4 revealed how many organisms overcome their need for the essential co-factor thiamin through the biosynthesis of the thiazole moiety. The use of NAD as a substrate allows for the organism to use a readily available metabolite that serves many other purposes in the cell. Interestingly, the hydroxymethyl pyrimidine moiety is also made through catalysis by Thi5 with pyridoxal 5´-phosphate (PLP) – another essential cofactor (*1*). With such a prodigious supply of substrate available, this ensures that the organism will be able to generate thiamin whenever it is needed. Structurally, Thi4 exhibits an example of an enzyme that retained a fold that was once able to bind the co-factor FAD, but evolved the ability to bind NAD as a substrate and no longer require a co-factor. This change also resulted in the inability of Thi4 to bind FAD. Still, many questions remain as to how Thi4 catalyzes NAD to ADT since this involves several catalytic steps. It has recently been discovered that the penultimate intermediate does not spontaneously aromatize into ADT, but rather requires a catalytic step to form the five-membered heterocyclic ring of thiazole (*submitted*). Also, the precise nature of the sulfur transfer step still remains elusive.

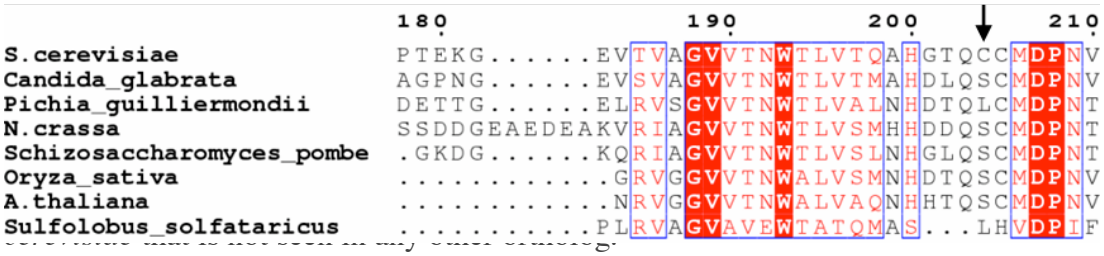
Originally it was believed that the sulfur atom from the thiazole ring came from a free cysteine amino acid. However, the partially active mutant of Thi4 (C204S) is able to generate advanced intermediates, but will not produce ADT if incubated in the presence of free cysteine (*2*). Also, recent crystallographic and biochemical evidence points to the possibility that the sulfur atom comes from an internal cysteine residue (Cys205) which is exclusively conserved in all Thi4 orthologs. The sulfur transfer would result in a dehydroalanine residue rendering the enzyme dead. X-ray electron density maps from *A. thaliana* (*3*) and *N. crassa* (unpublished data) Thi4 orthologs show no electron density where the γS atom would be located on this catalytic cysteine. A C205S mutant crystal structure of Thi4 from *S. cerevisiae* shows clear density for the γO atom, and is completely inactive (Appendix A Supplemental Figure 1). The di-cysteine motif seen in *S. cerevisiae* is unique to that organism as no other known sequence possesses an additional cysteine residue next to the catalytic cysteine (Appendix A Supplemental Figure 2). This leads to the speculation that the sulfur



Appendix A Supplemental Figure 1. Electron density maps highlighting the catalytic cysteine residue in Thi4. A) Simulated annealing composite omit map from *A. thaliana* (PDB accession code 1RP0). B) 16-fold averaged SA omit map from the Thi4 ortholog in *N. crassa*. C) Electron density map for the *S. cerevisiae* C205S mutant.

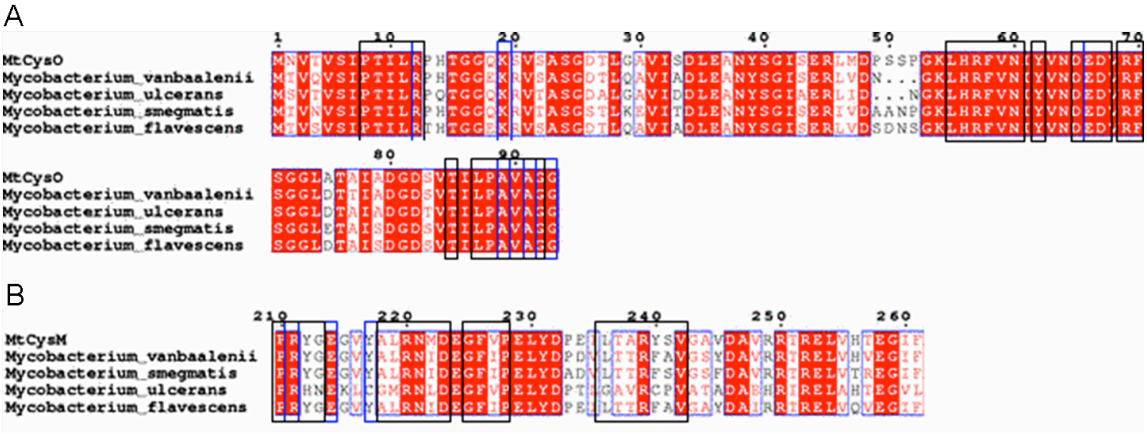
transfer step could be followed by the formation of a seven-membered heterocycle resulting from a nucleophilic substitution by the thiolate from Cys204 onto the Cβ carbon of the newly formed dehydroalanine. In the case of *N. crassa* and *A. thaliana*, the hydroxyl group of the adjacent serine residue is far less likely to act as a nucleophile, since its pKa is around 16 compared to a pKa of 8.3 for the thiol group on cysteine.

The structure of the CysM-CysO complex provides evidence for the mechanistic enzymology of cysteine biosynthesis from a structure based approach. Based on the orientation of the active site residues, it is clear that the Lys51, the same residue that forms a Schiff base with PLP, acts as the active site base to remove a proton from the Cα carbon of the O-acetylserine/PLP adduct. Previous structural studies of the aminoacrylate intermediate (*4*) and PLP/amino acid adducts (*5*) also led to the speculation that a structure based mechanism whereby the abstraction of the Cα carbon is done by the catalytic lysine residue responsible for forming the internal aldimine linkage to PLP. In the case of CysM-CysO, the mechanism does not involve sulfide as in the other examples. The use of a thiocarboxylated ubiquitin like protein to carry the sulfur to the α-aminoacrylate intermediate represents a clever example of evading the host's immune system by protecting the sulfur source from oxidation.



Furthermore, the CysM-CysO structure is additionally constrained in that the less reactive thiocarboxylate must be aligned with the intermediate so that the sulfur transfer may take place. In the afore mentioned examples, sulfide is the sulfur source which can easily move through a small channel in the protein to reach the intermediate. Sulfide has far fewer restrictions in movement than a protein like CysO, and therefore the CysM-CysO mechanism requires more structural constraints to achieve efficient catalysis.

Another interesting aspect of the CysM-CysO complex can be seen through sequence alignments between CysO and CysM orthologs. Alignments of CysO molecules from the Mycobacterium genus show that the residues which contribute to protein-protein interactions with CysM are conserved. Conversely, the same alignment with CysM orthologs shows that the CysO binding loop is *not* conserved. This unexpected difference implies that CysO, as well as other sulfur carrier proteins with a similar biological role, has the ability to bind to a diverse range of protein surfaces and



Appendix A Supplemental Figure 3. A) Sequence alignments between CysO orthologs. B) Sequence alignments between CysM orthologs that have a CysO binding partner. Residues involved in CysO-CysM binding are highlighted in boxes (black boxes indicate close contacts while blue boxes are involved in hydrogen bonding interactions).

may even be interchangeable between organisms.

The structure of PNP from *Thermotoga maritima* compared to bovine PNP is a classic example of key differences seen between thermophilic proteins and their mesophilic counterparts. No single cause has been attributed to the increased stability of thermophiles over mesophiles, yet all the characteristics that have been speculated to contribute to greater thermostability are seen in the TmPNP comparison with bPNP. One reason is that proteins able to withstand high temperatures without denaturing have more numerous and stronger intramolecular binding interactions. Typically thermophilic proteins have a greater number of salt bridges, π-cation interactions (*6*) hydrogen bonds and occasionally disulfide bridges (*7*) than their mesophilic counterparts. Additionally, characteristics attributed to overall structural rigidity are more prevalent in thermophiles than mesophiles. Some of these attributes include a greater number of proline residues, and a more compact structure with fewer flexible loop regions. Finally, chemical stability is realized by having fewer cysteine, asparagine and glutamine residues which can become oxidized at high temperatures (*8*).

Each of the above examples touches on active areas of structural biology research. Thi4 is an important contribution to the understanding of primary metabolism. CysM-CysO provides mechanistic insight into one of the myriad of defense mechanisms that *Mycobacterium tuberculosis* possesses against its host. Finally, the structure of TmPNP represents another example of the fundamentals behind protein stability. Together, what has been learned from these three systems represents a brief anthology to what the field of structural biology can contribute to science.

BIBLIOGRAPHY

(1) Wightman, R., and Meacock, P. A. (2003) The THI5 gene family of *Saccharomyces cerevisiae*: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine. *Microbiology 149*, 1447-60.

(2) Chatterjee, A., Jurgenson, C. T., Schroeder, F. C., Ealick, S. E., and Begley, T. P. (2007) Biosynthesis of thiamin thiazole in eukaryotes: conversion of NAD to an advanced intermediate. *J. Am. Chem. Soc. 129*, 2914-22.

(3) Godoi, P. H., Galhardo, R. S., Luche, D. D., Van Sluys, M. A., Menck, C. F., and Oliva, G. (2006) Structure of the thiazole biosynthetic enzyme THI1 from *Arabidopsis thaliana*. *J. Biol. Chem. 281*, 30957-66.

(4) Schnell, R., Oehlmann, W., Singh, M., and Schneider, G. (2007) Structural Insights into Catalysis and Inhibition of O-Acetylserine Sulfhydrylase from *Mycobacterium tuberculosis*: Crystal structures of the enzyme α-aminoacrylate intermediate and an enzyme-inhibitor complex. *J. Biol. Chem. 282*, 23473-23481.

(5) Rege, V. D., Kredich, N. M., Tai, C. H., Karsten, W. E., Schnackerz, K. D., and Cook, P. F. (1996) A change in the internal aldimine lysine (K42) in O-acetylserine sulfhydrylase to alanine indicates its importance in transimination and as a general base catalyst. *Biochemistry 35*, 13485-93.

(6) Chakravarty, S., and Varadarajan, R. (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry 41*, 8152-61.

(7) Zhang, Y., Porcelli, M., Cacciapuoti, G., and Ealick, S. E. (2006) The crystal structure of 5´-deoxy-5´-methylthioadenosine phosphorylase II from

Sulfolobus solfataricus, a thermophilic enzyme stabilized by intramolecular disulfide bonds. *J. Mol. Biol. 357*, 252-62.

(8) Tomazic, S. J., and Klibanov, A. M. (1988) Mechanisms of irreversible thermal inactivation of Bacillus α-amylases. *J. Biol. Chem. 263*, 3086-91.